Genetic mechanisms of critical illness in Covid-19

Genetic mechanisms of critical illness in Covid-19

Erola Pairo-Castineira*,[1,2] Sara Clohisey*,[1] Lucija Klaric*,[2] Andrew D. Bretherick*,[2] Konrad Rawlik*,[1] Dorota Pasko,[3] Susan Walker,[3] Nick Parkinson,[1] Max Head Fourman,[1] Clark D Russell,[1,4] James Furniss,[1] Anne Richmond,[2] Elvina Gountouna,[5] Nicola Wrobel,[6] David Harrison,[7] Bo Wang,[1] Yang Wu,[8] Alison Meynert,[2] Fiona Griffiths,[1] Wilna Oosthuyzen,[1] Athanasios Kousathanas,[3] Loukas Moutsianas,[3] Zhijian Yang,[9] Ranran Zhai,[9] Chenqing Zheng,[9] Graeme Grimes,[2] Rupert Beale,[10] Jonathan Millar,[1] Barbara Shih,[1] Sean Keating,[11] Marie Zechner,[1] Chris Haley,[1] David J. Porteous,[5] Caroline Hayward,[2,5] Jian Yang,[12,13] Julian Knight,[14] Charlotte Summers,[15] Manu Shankar-Hari,[16,17] Paul Klenerman,[14] Lance Turtle,[18] Antonia Ho,[19] Shona C Moore,[18] Charles Hinds,[20] Peter Horby,[21] Alistair Nichol,[22,23,24] David Maslove,[25] Lowell Ling,[26] Danny McAuley,[27,28] Hugh Montgomery,[29] Timothy Walsh,[11] Alex Pereira,[30] Alessandra Renieri,[31,32] The GenOMICC Investigators, The ISARIC-4C Investigators, The Covid-19 Human Genetics Initiative, 23andMe Investigators, BRACOVID Investigators, Gen-COVID Investigators, Xia Shen,[9,33,34] Chris P. Ponting,[2] Angie Fawkes,[6] Albert Tenesa,[1,2,33] Mark Caulfield,[3,20] Richard Scott,[3,35] Kathy Rowan,[7] Lee Murphy,[6] Peter J.M. Openshaw,[36] Malcolm G. Semple,[37] Andrew Law,[1] Veronique Vitart,[2] James F. Wilson,[2,33] J. Kenneth Baillie.[1,11]

Affiliations

1. Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, EH25 9RG, UK.
2. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK.
3. Genomics England, London, UK.
4. University of Edinburgh Centre for Inflammation Research, The Queen's Medical Research Institute, Edinburgh, UK.
5. Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK.
6. Edinburgh Clinical Research Facility, Western General Hospital, University of Edinburgh, EH4 2XU, UK.
7. Intensive Care National Audit & Research Centre, London, UK.
8. Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia
9. Biostatistics Group, School of Life Sciences, Sun Yat-sen University, Guangzhou, China.
10. The Crick Institute, London, UK
11. Intensive Care Unit, Royal Infirmary of Edinburgh, 54 Little France Drive, Edinburgh, EH16 5SA, UK.
12. School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China
13. Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China
14. Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.
15. Department of Medicine, University of Cambridge, Cambridge, UK.
16. Department of Intensive Care Medicine, Guy's and St. Thomas NHS Foundation Trust, London, UK.
17. School of Immunology and Microbial Sciences, King's College London, UK.
18. NIHR Health Protection Research Unit for Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences University of Liverpool, Liverpool, L69 7BE, UK.
19. MRC-University of Glasgow Centre for Virus Research, Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK.

20. William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK.

21. Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Old Road Campus, Roosevelt Drive, Oxford, OX3 7FZ, UK.

22. Clinical Research Centre at St Vincent's University Hospital, University College Dublin, Dublin, Ireland.

23. Australian and New Zealand Intensive Care Research Centre, Monash University, Melbourne, Australia.

24. Intensive Care Unit, Alfred Hospital, Melbourne, Australia.

25. Department of Critical Care Medicine, Queen's University and Kingston Health Sciences Centre, Kingston, ON, Canada.

26. Department of Anaesthesia and Intensive Care, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China.

27. Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland, UK.

28. Department of Intensive Care Medicine, Royal Victoria Hospital, Belfast, Northern Ireland, UK.

29. UCL Centre for Human Health and Performance, London, W1T 7HA, UK.

30. BRACOVID Study

31. Medical Genetics, University of Siena, Italy

32. Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Italy

33. Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, Teviot Place, Edinburgh EH8 9AG, UK.

34. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

35. Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK.

36. National Heart & Lung Institute, Imperial College London (St Mary's Campus), Norfolk Place, Paddington, London W2 1PG, UK.

37. University of Liverpool, Liverpool, UK.

* Joint first authorship. These authors contributed equally to this work.

Address for correspondence: Dr J Kenneth Baillie, Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, EH25 9RG, UK.

email: j.k.baillie@ed.ac.uk

Host-mediated lung inflammation is present,[1] and drives mortality,[2] in critical illness caused by Covid-19. Host genetic variants associated with critical illness may identify mechanistic targets for therapeutic development.[3]

Here we report the results of the GenOMICC (Genetics Of Mortality In Critical Care) genome-wide association study(GWAS) in 2244 critically ill Covid-19 patients from 208 UK intensive care units (ICUs).

We identify and replicate novel genome-wide significant associations, on chr12q24.13 (rs10735079, p=1.65 $\times$ 10^{-8}) in a gene cluster encoding antiviral restriction enzyme activators (*OAS1*, *OAS2*, *OAS3*), on chr19p13.2 (rs2109069, p=2.3 $\times$ 10^{-12}), near the gene encoding tyrosine kinase 2 (*TYK2*), on chr19p13.3 (rs2109069, p=3.98 $\times$ 10^{-12}), within the gene encoding dipeptidyl peptidase 9 (*DPP9*), and on chr21q22.1 (rs2236757, p=4.99 $\times$ 10^{-8}) in the interferon receptor gene *IFNAR2*.

We identify potential targets for repurposing of licensed medications: using Mendelian randomisation we found evidence in support of a causal link from low expression of *IFNAR2*, and high expression of *TYK2*, to life-threatening disease; transcriptome-wide association in lung tissue revealed that high expression of the monocyte/macrophage chemotactic receptor *CCR2* is associated with severe Covid-19.

Our results identify robust genetic signals relating to key host antiviral defence mechanisms, and mediators of inflammatory organ damage in Covid-19. Both mechanisms may be amenable to targeted treatment with existing drugs. Large-scale randomised clinical trials will be essential before any change to clinical practice.

Since critical illness in Covid-19 is caused, in part, by inflammatory injury affecting the lungs and lung blood vessels.[4,5], there are at least two distinct biological components to mortality risk: susceptibility to viral infection, and propensity to develop harmful pulmonary inflammation. Susceptibility to life-threatening infections[6] and immune-mediated diseases are both strongly heritable. In particular, susceptibility to respiratory viruses[7] such as influenza[8] is heritable and known to be associated with specific genetic variants.[9] In Covid-19, one genetic locus, 3p21.31, has been repeatedly associated with hospitalisation.[10,11] As with other viral illnesses,[12] there are several examples of loss-of-function variants affecting essential immune processes that lead to severe disease in young people: for example *TLR7*,[13], and several genes implicated in type 1 interferon signalling including the receptor subunit *IFNAR2*.[14] Genome-wide studies have the potential to reveal completely new molecular mechanisms of critical illness in Covid-19, which may provide therapeutic targets to modulate the host immune response to promote survival.[3]

There is now strong evidence that critical illness caused by Covid-19 is qualitatively different from mild or moderate disease, even among hospitalised patients. There are multiple distinct disease phenotypes with differing patterns of presenting symptoms[15] and marked differential responses to immunosuppressive therapy.[2] In patients without respiratory failure, there is a trend towards harm from treatment with corticosteroids, whereas among patients with critical respiratory failure, there is a very substantial benefit.[2] On this basis, we consider patients with critical Covid-19 respiratory failure to have distinct pathophysiology.

In the UK, the group of patients admitted to critical care is relatively homogeneous, with profound hypoxaemic respiratory failure being the archetypal presentation.[16] The active disease process in these patients is strikingly responsive to corticosteroid therapy[17] and is characterised by pulmonary inflammation including diffuse alveolar damage, lung macrophage/monocyte influx, mononuclear cell pulmonary artery vasculitis and microthrombus formation.[4,5]

Host-directed therapies have long been an aspiration for the treatment of severe disease caused by respiratory viruses.[18] Identification of genetic loci associated with susceptibility to Covid-19 may lead to specific targets for repurposing or drug development.[3]

The GenOMICC (Genetics Of Mortality In Critical Care, genomicc.org) study has been recruiting patients with critical illness syndromes, including influenza, sepsis, and emerging infections, for 5 years. In order to better understand the host mechanisms leading to life-threatening Covid-19, we performed a genome-wide

association study comparing critically ill patients with Covid-19 with controls from two population genetic studies in the UK.

Results

Critically ill cases were recruited through the GenOMICC study in 208 UK Intensive Care Units and hospitalised cases through the International Severe Acute Respiratory Infection Consortium (ISARIC) Coronavirus Clinical Characterisation Consortium (4C) study. Cases were representative of the UK critically ill population.[16] Demographic and summary clinical characteristics of ICU recruited participants analysed in the GWAS are described in Supplementary Table 2.

DNA was extracted from whole blood and array-based genome-wide genotypes of good quality obtained for 2734 unique individuals (Materials & Methods). Genetic ancestry was inferred using principal component analyses and individuals from the 1000 Genomes project as population references (Materials & Methods). After quality control and matching to ancestry groups, 2244 individuals were included for GWAS analysis. Clinical and demographic features of these cases are shown in Supplementary Table 2. Additional clinical details for a subset of 1069 cases for whom additional data was available is presented in Supplementary Figures 8-13. Imputation in this multi-ancestry cohort was performed using the TOPMed reference panel.[19].

Ancestry-matched controls were selected from the large population-based cohort UK Biobank in a ratio of 5 controls to 1 case. Controls with a known positive Covid-19 test were excluded. The inevitable presence of individuals in the control group, who may exhibit the critical illness phenotype if exposed to SARS-CoV-2, is expected to bias any associations towards the null. GWAS was carried out separately by ancestry group using logistic regression in PLINK and accounting for age, sex, postal code deprivation decile and principal components of ancestry. As well as several standard filters to minimise spurious associations (Materials & Methods), whole genome sequencing of a subset of 1613 cases was used to filter out variants likely to have been badly called or imputed; 83937 out of the 4469187 imputed variants that passed other quality control filters after GWAS were thus removed. There was a high level of residual inflation in the South Asian and East Asian ancestry groups, rendering results in these subgroups unreliable (Supplementary Figure 5). The largest ancestry group contained 1676 individuals of European descent (EUR); this group were used for the primary analyses presented below.

GWAS results

In the primary analysis (GenOMICC European cases vs. UK Biobank controls), following linkage disequilibrium-based clumping, 15 independent association signals were genome-wide significant at $p < 5 \times 10^{-8}$ (Figure 1). Eight of these were successfully validated in GWAS using two independent population genetic studies (100,000 genomes and Generation Scotland) as controls (Table 1) and hence were taken forward for replication. A sex-specific GWAS among this group found no sex-specific associations (Supplementary Table 1). Trans-ethnic meta-analysis did not reveal additional associations (Supplementary Figure 2).

Table 1: Lead variants from independent genome-wide significant regions. chr:pos - chromosome and position of the top SNP (build 37); Risk – risk allele; Alt - other allele; RAF - risk allele frequency; OR - effect size (odds ratio) of the risk allele in the GenOMICC EUR analysis; CI - 95% confidence interval for the odds ratio in the GenOMICC EUR cohort; P - p-value, Locus – gene nearest to the top SNP. Subscript identifiers indicate the cohorts used for cases: gcc - GenOMICC EUR; and controls: ukb - UK Biobank; gs - Generation Scotland; 100k - 100,000 genomes.

SNP	chr:pos(b37)	Risk	Alt	RAF_{gcc}	RAF_{ukb}	OR	CI	$P_{gcc.ukb}$	$P_{gcc.gs}$	$P_{gcc.100k}$	Locus
rs73064425	3:45901089	T	C	0.15	0.07	2.1	1.88-2.45	4.8×10^{-30}	2.9×10^{-27}	3.6×10^{-32}	*LZTFL1*
rs9380142	6:29798794	A	G	0.74	0.69	1.3	1.18-1.43	3.2×10^{-8}	0.00091	1.8×10^{-8}	*HLA-G*
rs143334143	6:31121426	A	G	0.12	0.07	1.9	1.61-2.13	8.8×10^{-18}	2.6×10^{-24}	5.8×10^{-18}	*CCHCR1*
rs3131294	6:32180146	G	A	0.9	0.86	1.5	1.28-1.66	2.8×10^{-8}	1.3×10^{-10}	2.3×10^{-8}	*NOTCH4*
rs10735079	12:113380008	A	G	0.68	0.63	1.3	1.18-1.42	1.6×10^{-8}	2.8×10^{-9}	4.7×10^{-6}	*OAS1/3*
rs2109069	19:4719443	A	G	0.38	0.32	1.4	1.25-1.48	4×10^{-12}	4.5×10^{-7}	2.4×10^{-8}	*DPP9*
rs74956615	19:10427721	A	T	0.079	0.05	1.6	1.35-1.87	2.3×10^{-8}	2.2×10^{-13}	3.9×10^{-6}	*TYK2*

SNP	chr:pos(b37)	Risk	Alt	RAF_{gcc}	RAF_{ukb}	OR	CI	$P_{gcc.ukb}$	$P_{gcc.gs}$	$P_{gcc.100k}$	Locus
rs2236757	21:34624917	A	G	0.34	0.28	1.3	1.17-1.41	5×10^{-8}	8.9×10^{-5}	8.3×10^{-7}	*IFNAR2*

Replication

Since no study of critical illness in Covid-19 of sufficient size is available, replication was sought in a meta-analysis of data from 2415 hospitalised Covid-19 cases and 477741 population controls from the Covid-19 Host Genetics Initiative (HGI, mixed ancestry, with UK Biobank cases and controls excluded) and 1128 cases and 679531 controls in the 23andMe Inc "broad respiratory phenotype" (EUR ancestry), which includes cases reported being placed on a ventilator, being administered oxygen, or having pneumonia versus controls who did not report positive tests. In addition to the locus on chr3 already reported (rs73064425, OR=2.14, discovery $p=4.77 \times 10^{-30}$), we found robust replication for the novel associations in four loci from GenOMICC: a locus on chr12 in the *OAS* gene cluster (rs74956615, OR=1.59, discovery $p = 1.65 \times 10^{-8}$), near *TYK2* on chr19 (rs74956615, OR=1.4, discovery $p = 2.3 \times 10^{-8}$), in *DPP9* on chr19 (rs2109069, OR=1.36, discovery $p = 3.98 \times 10^{-12}$), and a locus on chromosome 21, containing the gene *IFNAR2* (rs2236757, OR=1.28, discovery $p = 4.99 \times 10^{-8}$) (Table 2, Figure 2).

Three variants, all in a region of the genome in which population stratification is difficult to control (the major histocompatibility complex, MHC), did not replicate (Table 2). Further studies will be required to determine whether these associations are real.

Table 2: Replication in external data from Covid-19 HGI study. Risk – risk allele; Alt - alternative allele; OR - effect size (odds ratio) of the risk allele; CI - 95% confidence interval for the odds ratio; P - p-value, locus – gene nearest to the top SNP. Subscript identifiers show the data source: gcc - GenOMICC study, European ancestry, comparison with UK Biobank; hgi.23m - Covid-19 human genetics initiative and 23andMe meta-analysis, used for replication. * Bonferroni significant values are highlighted and indicate external replication.

SNP	chr:pos(b37)	Risk	Alt	$OR_{gcc.ukb}$	$P_{gcc.ukb}$	$OR_{hgi.23m}$	$P_{hgi.23m}$	Locus
rs73064425	3:45901089	T	C	2.1	4.8×10^{-30}	1.7	1.5×10^{-28}*	*LZTFL1*
rs9380142	6:29798794	A	G	1.3	3.2×10^{-8}	1	0.76	*HLA-G*
rs143334143	6:31121426	A	G	1.9	8.8×10^{-18}	1.1	0.019	*CCHCR1*
rs3131294	6:32180146	G	A	1.5	2.8×10^{-8}	0.99	0.91	*NOTCH4*
rs10735079	12:113380008	A	G	1.3	1.6×10^{-8}	1.1	0.00082*	*OAS1/3*
rs2109069	19:4719443	A	G	1.4	4×10^{-12}	1.1	5×10^{-5}*	*DPP9*
rs74956615	19:10427721	A	T	1.6	2.3×10^{-8}	1.4	2×10^{-6}*	*TYK2*
rs2236757	21:34624917	A	G	1.3	5×10^{-8}	1.2	4.1×10^{-5}*	*IFNAR2*

To further increase power for exploratory downstream analyses meta-analysis was performed using inverse-variance meta-analysis,[20] between GenOMICC critically ill EUR ($n_{cases} = 1676$, $n_{controls} = 8380$), HGI hospitalised Covid-19 vs population (B2, version 2) without UKBioBank ($n_{cases} = 2415$, $n_{controls} = 477741$) and the 23andMe broad respiratory phenotype ($n_{cases} = 1128$, $n_{controls} = 679531$). This revealed one additional (unreplicated) locus in CCHCR1 at genome-wide significance (using a more stringent threshold of $p<10^{-8}$ in view of the absence of replication opportunities for the meta-analysis).

Table 3: Meta-analysis of overlapping SNPs between GenOMICC (EUR) and HGI (hospitalized Covid-19 vs. population) and 23andMe studies. Since this is a meta-analysis of all available data, external replication cannot be attempted, so SNPs are included in this table if they meet a more stringent p-value threshold of $p<10^{-8}$. SNP – the strongest SNP in the locus, ; Risk – risk allele; Alt - alternative allele; OR - odds ratio of the risk allele; CI - 95% confidence interval for odds ratio; Locus – gene nearest to the top SNP. Subscript identifiers show gcc - GenoMICC study, European ancestry, comparison with UK Biobank; meta - combined meta-analysis of all three studies (GenOMICC, HGI and 23andMe) for cases of European ancestry.

SNP	chr:pos(b37)	Risk	Alt	OR_{gcc}	CI_{gcc}	P_{gcc}	OR_{meta}	CI_{meta}	P_{meta}	Locus
rs71325088	3:45862952	C	T	2.1	1.87-2.43	9.3×10^{-30}	1.9	1.73-2	2.5×10^{-54}	*LZTFL1*
rs143334143	6:31121426	A	G	1.8	1.61-2.13	8.8×10^{-18}	1.3	1.27-1.48	1.5×10^{-10}	*CCHCR1*

SNP	chr:pos(b37)	Risk	Alt	OR_{gcc}	CI_{gcc}	P_{gcc}	OR_{meta}	CI_{meta}	P_{meta}	Locus
rs6489867	12:113363550	T	C	1.3	1.15-1.37	6.9×10^{-7}	1.2	1.14-1.25	9.7×10^{-10}	*OAS1*
rs2109069	19:4719443	A	G	1.4	1.25-1.48	4×10^{-12}	1.2	1.19-1.31	7×10^{-13}	*DPP9*
rs11085727	19:10466123	T	C	1.3	1.17-1.4	1.3×10^{-7}	1.2	1.18-1.31	1.2×10^{-13}	*TYK2*
rs13050728	21:34615210	T	C	1.3	1.15-1.38	3×10^{-7}	1.2	1.16-1.28	5.1×10^{-12}	*IFNAR2*

Mendelian randomisation

Mendelian randomisation provides evidence for a causal relationship between an exposure variable and an outcome, given a set of well-characterised assumptions.[21] We employ two-sample summary-data Mendelian randomisation[22] here to assess the evidence in support of causal effects of RNA expression (GTEx v7, whole blood) of various genes on the odds of critical Covid-19.

We specified an *a priori* list of target genes that relate to the mechanism of action of many host-targeted drugs that have been proposed for the treatment of Covid-19 (Supplementary Table 4). Seven of these targets had a suitable locally acting expression quantitative trait locus (eQTL) in GTEx(v7). Of these, *IFNAR2* remained significant after Bonferroni correcting for multiple testing for 7 tests (β -1.49, standard error 0.52, p = 0.0043). There was equivocal evidence of heterogeneity (HEIDI[22] p = 0.015), indicating that the effect of this variant on critical illness in Covid-19 may be mediated through another mechanism, which may lead to an under- or over-estimation of the effect of IFNAR2 expression on risk of critical illness.

We then performed transcriptome-wide Mendelian randomisation to quantify support for *unselected* genes as potential therapeutic targets. Instruments were available for 4,614 unique Ensembl gene IDs. No genes were statistically significant after correcting for multiple comparisons in this analysis (4,614 tests). After conservative filtering for heterogeneity (HEIDI $p > 0.05$), the smallest Mendelian randomisation p = 0.00049 for a variant at chr19:10466123 affecting expression of *TYK2*. 9 other genes with nominally significant Mendelian randomisation p-values ($p<0.0051$) were also taken forward for further analysis.

To replicate these findings, we tested for external evidence using a separate eQTL dataset (eQTLgen)[23] and GWAS (HGI B2, excluding UK Biobank). Mendelian randomisation signals with consitent directions of effect were significant for *IFNAR2* ($p = 7.5 \times 10^{-4}$) and *TYK2* ($p = 5.5 \times 10^{-5}$).

Transcriptome-wide association study

We performed a transcriptome-wide association study(TWAS)[24,25] to link GWAS results to tissue-specific gene expression data by inferring gene expression from known genetic variants that are associated with transcript abundance (eQTL). For this analysis we used GTEx v8 data for two disease-relevant tissues chosen *a priori*: whole blood and lung(Figure 3). We selected genes with $p<0.05$ in these tissues and performed a combined meta-TWAS analysis,[26] incorporating eQTL data from other tissues in GTEX, to optimise power to detect differences in predicted expression in lung or blood.

We discovered 5 genes with genome-wide significant differences in predicted expression compared to controls (Supplementary Table 8). This included 4 genes with differential predicted expression in lung tissue (Figure 3; 3 on chr3: *CCR2*, *CCR3* and *CXCR6*, and one on chr5: *MTA2B*).

We used meta-analysis by information content (MAIC)[27] to put these results in the context of existing biological knowledge about host-virus interactions in Covid. We combined the top 2000 metaTWAS genes with previous systematically compiled experimental evidence implicating human genes in SARS-CoV-2 replication and host response. MAIC derives a data-driven weighting for each of a range of experimental data sources in the form of gene lists, and outperforms other approaches to providing a composite of multiple lists.[27] We found that the GenOMICC TWAS results had greater overlap with results from transcriptomic, proteomic and CRISPR studies of host genes implicated in Covid-19 than any other data source(Supplementary Figure 22).

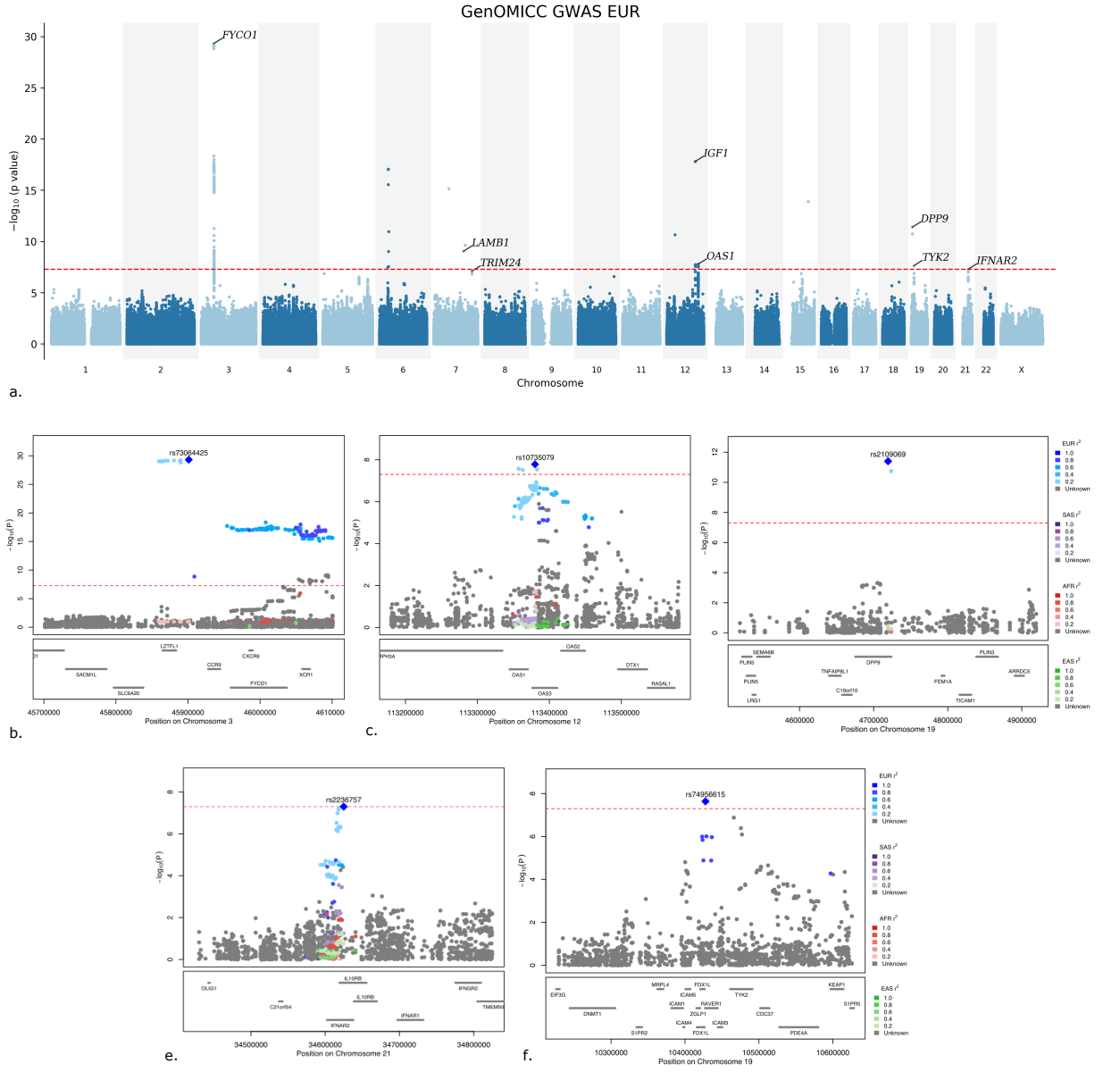

Figure 1: Summary of GWAS results for EUR ancestry group in GenOMICC. *a.* Manhattan plot showing single nucleotide polymorphism (SNP)-level p-values for genome-wide significant associations in largest ancestry group, EUR (red horizontal line shows genome-wide significance at $-log_{10}(5 \times 10^{-8})$) *b-d.* Genomic region plots showing ancestry-specific p-values and LD structure. Each variant is plotted at the y-axis position for GWAS within a given ancestry group in GenOMICC vs UK Biobank analysis. A variant is plotted for an ancestry group only if it has MAF>5% in that group in UK Biobank. Colour depicts genetic ancestry group (EUR = blue, EAS = green, AFR = red, SAS = purple); shading shows linkage disequilibrium value (LD, r2) within a given ancestry group with the lead SNP; variants with unknown LD for each ancestry group are shown in grey; lead SNP for EUR is shown in each plot as a blue diamond. LD reference is calculated using PLINK from 5 cases randomly selected unrelated matched-ethnicity subjects in UK Biobank.

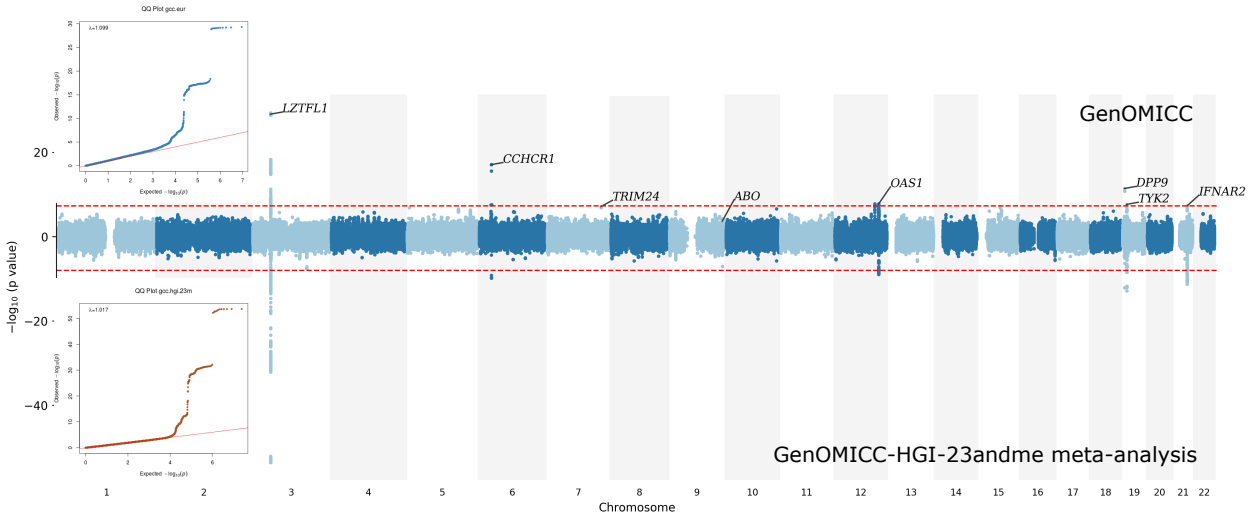

Figure 2: Miami plot showing relationship between the GenOMICC GWAS and meta-analysis including patients from the Covid-19 Human Genetics Initiative and 23andMe. In upper (GenOMICC) panel, red horizontal line shows genome-wide significance for common variants at $-log_{10}(5 \times 10^{-8})$; in lower (meta-analysis) panel, red horizontal line shows a more stringent genome-wide significance threshold for meta-analysis variants at $-log_{10}(10^{-8})$. Quantile-quantile (QQ) plots are inset showing genomic inflation (λ) for each analysis: GenOMICC EUR $\lambda = 1.099$; GenOMICC-HGI-23m meta-analysis $\lambda = 1.017$

Genetic correlations and tissue associations

We used the high-definition likelihood (HDL) method[28] to provide an initial estimate the SNP-based heritability, that is the proportion of phenotypic variance that is captured by additive effects at common SNPs, to be 0.065 (SE = 0.019) for severe Covid-19. We were not able to detect a significant signal for heritability in two additional analyses: firstly, using controls from the 100,000 genomes project (in which matching to the GenOMICC cases is less close, which may limit heritability estimation) and secondly, in a smaller GWAS comparing some GenOMICC cases with UK Biobank controls, using matching of BMI and age where possible. This second analysis was less powerful because of the lack of close matches for many cases ($n_{cases} = 1260$; $n_{controls} = 6300$; Supplementary Figure 15). Including rare variants in future analyses, with larger numbers of cases, will provide a more comprehensive estimate of heritability. We also tested for genetic correlations with other traits, that is, the degree to which the underlying genetic components are shared with severe Covid-19. Using the HDL method, we identified significant negative genetic correlations with educational attainment and intelligence. Significant positive genetic correlations were detected for a number of adiposity phenotypes including body mass index and leg fat, as well as pulse rate, neck/shoulder pain, and treatment with thyroxine and lansoprazole (Supplementary Figure 20).

Consistent with GWAS results from other infectious and inflammatory diseases,[29] there was a significant enrichment of strongly associated variants in promoters and enhancers, particularly those identified by the EXaC study as under strong evolutionary selection (Supplementary Figure 19).[30] The strongest tissue type enrichment was in spleen, which may reflect enrichment in immune cells, followed by pancreas (Supplementary Figure 21).

Discussion

We have discovered and replicated significant genetic associations with life-threatening Covid-19 (Figure 1). Our focus on critical illness increases the probability that some of these associations relate to the later, immune-mediated phase of disease associated with respiratory failure requiring invasive mechanical ventilation.[2] Importantly, the GWAS approach is unbiased and genome-wide, enabling the discovery of completely new pathophysiological mechanisms. Because genetic variation can be used to draw a causal inference,

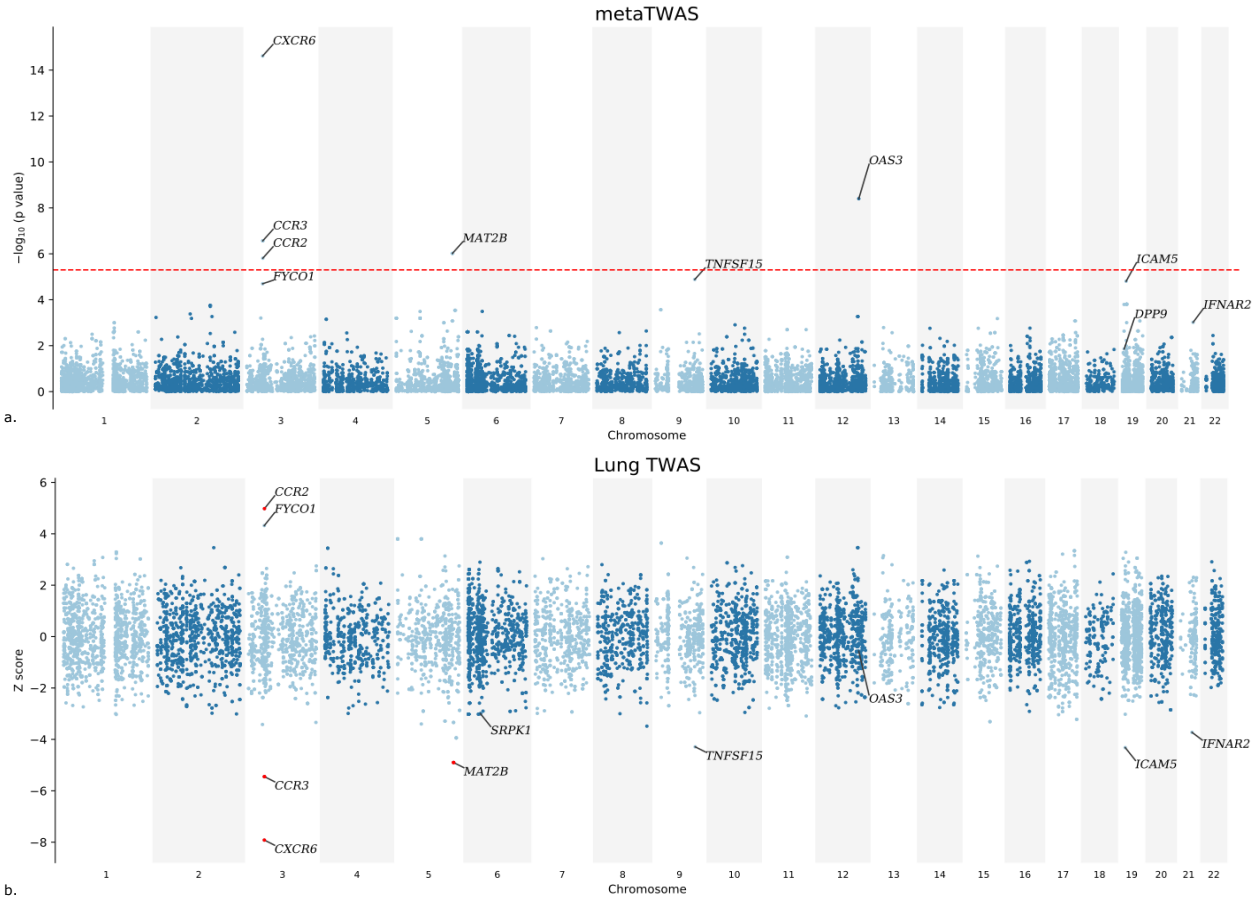

Figure 3: Summary of TWAS results. *a.* Gene-level Manhattan plot showing results from meta-TWAS analysis across tissues. Red horizontal line shows gene-level genome-wide significance at $-log_{10}(5 \times 10^{-6})$ *b.* z-scores showing direction of effect for genotype-inferred expression of transcripts encoding protein-coding genes in lung tissue (GTEX v8). Red highlighting indicates genome-wide significance at $p < 5 \times 10^{-6}$.

genetic evidence in support of a therapeutic target substantially improves the probability of successful drug development.[31] In particular, Mendelian randomisation occupies a unique position in the hierarchy of clinical evidence.[32]

Patients admitted to intensive care units in the UK during the first wave of Covid-19 were, on average, younger and less burdened by comorbid illness than the hospitalised population.[16] The population studied here are defined by their propensity to critical respiratory failure due to Covid-19. GenOMICC recruited in 208 intensive care units (covering > 95% of UK ICU capacity), ensuring that a broad spread across the genetic ancestry of UK patients was included (Supplementary Figure 3).

For external replication, the nearest comparison is the hospitalised vs population analysis in the Covid-19 Host Genetics initiative, and the 23andMe broad respiratory phenotype, which have been generously shared with the international community. Likewise, full summary statistics from GenOMICC have been made immediately openly available at genomicc.org/data.

Despite the differences in case definitions, novel associations from our study of critical illness replicate robustly in combined data from hospitalised case studies (Table 2). Separately, the Mendelian randomisation results implying a causal role for *IFNAR2* and *TYK2* are also statistically significant in confirmatory analyses. Our findings reveal that critical illness in Covid-19 is related to at least two biological mechanisms: innate antiviral defences, which are known to be important early in disease (*IFNAR2* and *OAS* genes), and host-driven inflammatory lung injury, which is a key mechanism of late, life-threatening Covid-19 (*DPP9*, *TYK2* and *CCR2*).[2]

Interferons are canonical host antiviral signalling mediators, and stimulate release of many essential components of the early host response to viral infection.[33] Consistent with a beneficial role for type I interferons, increased expression of the interferon receptor subunit *IFNAR2* reduced the odds of severe Covid-19 with Mendelian randomisation discovery $p = 0.0043$ (7 tests); replication $p = 7.5 \times 10^{-4}$ (1 test). Within the assumptions of Mendelian randomisation, this represents evidence for a protective role for IFNAR2 in Covid-19. Rare loss-of-function mutations in *IFNAR2* are associated with severe Covid-19[14] and many other viral diseases.[34,35] This suggests that adminstration of interferon may reduce the probability of critical illness in Covid-19, but our evidence cannot inform when in illness such a treatment may be effective. Exogenous interferon treatment did not reduce mortality in hospitalised patients in a large scale clinical trial,[36] suggesting that this genetic effect may be mediated during the early phase of disease when viral load is high.

The variant rs10735079 (chr12, $p = 1.65 \times 10^{-8}$) lies in the oligoadenylate synthetase (OAS) gene cluster (*OAS1*, *OAS2* and *OAS3*; Figure 1), which are inducibly expressed in respose to Type 1 interferon signalling. TWAS also (Figure 3) detects significant associations with predicted expression of OAS genes (Figure 3), and *OAS1* variants were implicated in susceptibility to SARS-CoV in candidate gene association studies in Vietnam[37] and China.[38] These genes encode enzymes which produce a mediator (2',5'-oligoadenylate, 2-5A) which activates an effector enzyme, RNAse L. RNAse L degrades double-stranded RNA,[39] a replication intermediate of coronaviruses.[40] The OAS genes therefore also provide a potential therapeutic target, since the antiviral consequences of OAS genes are counteracted by endogenous phosphodiesterase 12 (PDE-12) activity, which degrades the mediator 2-5A; inhibiting PDE-12 augments OAS-mediated antiviral activity.[41] The betacoronaviruses OC43 and MHV make viral phosphodiesterases that cleave the host antiviral mediator 2-5A,[42] but SARS-CoV-2 is not known to have this ability.

The association in 19p13.3 (rs2109069, $p = 3.98 \times 10^{12}$) is an intronic variant in the gene encoding dipeptidyl peptidase 9 (*DPP9*). Variants in this locus are associated with idiopathic pulmonary fibrosis[43] and interstitial lung disease.[44] *DPP9* encodes a serSadler, Anthony J., and Bryan R. G. Williams. 'Interferon-Inducible Antiviral Effectors'. Nature Reviews Immunology 8, no. 7 (July 2008): 559–68. https://doi.org/10.1038/nri2314. ine protease with diverse intracellular functions, including cleavage of the key antiviral signalling mediator CXCL10,[45] key roles in antigen presentation,[46] and in inflammosome activation.[47]

Since opportunities for therapeutic intervention, particularly experimental therapy, are more abundant in later, more severe disease, it is important that our results also reveal genes that may act to drive inflammatory organ injury. *TYK2* is one of 4 gene targets for baricitinib,[48] one of the nine candidate drugs we used in the creation of our *a priori* target list (Supplementary Table 4). The association between *TYK2* expression and critical illness was also confirmed in an external dataset.

We replicate the finding of Ellinghaus *et al.* at 3p21.31.[11] The extremely small p-value at this locus (p=4.77 $\times$ 10^{-30}) may reflect the large size of our study, and our focus on extreme severity, since we see a greater effect size in GenOMICC than in the replication studies (Supplementary Figure 6). The 3p21.31 locus is populated by a number of genes with mechanisms of action that could plausibly explain an association. Our systematic review and meta-analysis of experimental data on betacoronavirus infection from other sources provides moderate biological support for *FYCO1*, although this additional information comes mostly from *in vitro* model systems.[49]

Our TWAS results show that variants in this region confer genome-wide significant differences in predicted expression of *CXCR6*, *CCR2* and *CCR3* (Figure 3 a); it is likely that one, but not all of these genes is an important mediator of critical illness. Association with critical illness for genotype-inferred *CCR2* (CC-chemokine receptor 2) expression is particularly strong in lung tissue(Figure 3 b). CCR2 promotes monocyte/macrophage chemotaxis towards sites of inflammation, and there is increased expression of the canonical ligand for CCR2 (monocyte chemoattractant protein/MCP-1), in bronchoalveolar lavage fluid from the lungs of Covid-19 patients during mechanical ventilation.[50] Circulating MCP-1 concentrations are associated with more severe disease.[51] Anti-CCR2 monoclonal antibody therapy in treatment of rheumatoid arthritis is safe.[52]

The *ABO* locus was also previously associated with Covid-19,[11] but was not genome-wide significant in the GenOMICC critically ill cohort. Interestingly there is a signal close to genome-wide significance at this locus in the combined meta-analysis (Figure 2), suggesting that this variant may be associated with susceptibility to Covid-19, but not critical illness (Supplementary Figure 6).

Analysis of shared heritability highlights a positive correlation with adiposity. This does not imply a causal relationship, as a number of biases may be at play, but may reflect a combination of two effects: firstly, increased BMI and lower socio-economic status are strong risk factors for severe Covid-19,[53,54] and secondly, UK Biobank participants are disproportionately drawn from social groups in which obesity is under-represented compared to the general population.[55]

Because of the urgency of completing and reporting this work, we have drawn controls from population genetic studies with systematic differences in population structure, demographics and comorbid illness, who were genotyped using different technology from the cases. Residual confounding is reflected in the genomic inflation ($\lambda_{0.5}$) value of 1.099 for the primary analysis (Figure 1). We mitigated the consequent risk of false-positive associations driven by genotyping errors by genotyping the majority of our subjects using two different methods (microarray and whole-genome sequencing), and by verifying significant associations using two additional control groups (100,000 genomes and Generation Scotland). The success of these mitigations is demonstrated by robust replication of our sentinel SNPs in external studies. Our meta-analysis, combining GenOMICC with multiple additional sources of genome-wide associations, has a reassuring $\lambda_{0.5} = 1.017$ (Figure 1).

There is an urgent need to deepen these findings through further studies. Our MAIC results show that highly ranked genes in GenOMICC are more likely to be implicated in Covid in other studies (Supplementary Figure 22). We continue to recruit to the GenOMICC study, in the expectation that additional associations exist and can be detected with larger numbers of cases. Future studies using whole genome sequencing will search the rarer end of the allele frequency spectrum for variants increasing susceptibility. Effect sizes are likely to be higher in GenOMICC because the cohort is strongly enriched for immediately life-threatening disease in patients who are either receiving invasive mechanical ventilation, or considered by the treating physicians to be at high risk of requiring mechanical support.

We have discovered new and highly plausible genetic associations with critical illness in Covid-19. Some of these associations lead directly to potential therapeutic approaches to augment interferon signalling, antagonise monocyte activation and infiltration into the lungs, or specifically target harmful inflammatory pathways. While this adds substantially to the biological rationale underpinning specific therapeutic approaches, each treatment must be tested in large-scale clinical trials before entering clinical practice.

Materials and methods

Recruitment of cases

2,636 patients recruited to the GenOMICC study (genomicc.org) had confirmed Covid-19 according to local clinical testing and were deemed, in the view of the treating clinician, to require continuous cardiorespiratory monitoring. In UK practice this kind of monitoring is undertaken in high-dependency or intensive care units. An additional 135 patients were recruited through ISARIC 4C (isaric4c.net) - these individuals had confirmed Covid-19 according to local clinical testing and were deemed to require hospital admission. Both studies were approved by the appropriate research ethics committees (Scotland 15/SS/0110, England, Wales and Northern Ireland: 19/WM/0247). Current and previous versions of the study protocol are available at genomicc.org/protocol.

Genotyping

DNA was extracted from whole blood using Nucleon Kit (Cytiva) with the BACC3 protocol. DNA samples were re-suspended in 1 ml TE buffer pH 7.5 (10mM Tris-Cl pH 7.5, 1mM EDTA pH 8.0). The yield of the DNA was measured using Qubit and normalised to 50ng/μl before genotyping.

Genotyping was performed using the Illumina Global Screening Array v3.0 + multi-disease beadchips (GSAMD-24v3-0-EA) and Infinium chemistry. In summary this consists of three steps: (1) whole genome amplification, (2) fragmentation followed by hybridisation, and (3) single-base extension and staining. For each of the samples, 4 μl of DNA normalised to 50ng/μl was used. Each sample was interrogated on the arrays against 730,059 SNPs. The arrays were imaged on an Illumina iScan platform and genotypes were called automatically using GenomeStudio Analysis software v2.0.3, GSAMD-24v3-0-EA_20034606_A1.bpm manifest and cluster file provided by manufacturer.

In 1667 cases, genotypes and imputed variants were confirmed with Illumina NovaSeq 6000 whole genome sequencing. Samples were aligned to the human reference genome hg38 and variant called to GVCF stage on the DRAGEN pipeline (software v01.011.269.3.2.22, hardware v01.011.269) at Genomics England. Variants were genotyped with the GATK GenotypeGVCFs tool v4.1.8.1,[56] filtered to minimum depth 8X (95% sensitivity for heterozygous variant detection,[57]) merged and annotated with allele frequency with bcftools v1.10.2.

Quality control

Genotype calls were carefully examined within GenomeStudio using manufacturer and published[58] recommendations, after excluding samples with low initial call rate (<90%) and reclustering the data thereafter. Briefly, X and Y marker calls were all visually inspected and curated if necessary, as were those for autosomal markers with minor allele frequency > 1% displaying low Gentrain score, cluster separation, and excess or deficit of heterozygous calls. Genotype-based sex determination was performed in GenomeStudio and samples excluded if not matching records expectation. Five individuals with XXY genotypes were also detected and excluded for downstream GWAS analyses. Genotypes were exported, in genome reference consortium human build 37 (GRCHb37) and Illumina "source" strand orientation, using the GenotypeStudio `plink-input-report-plugin-v2-1-4`. A series of filtering steps was then applied using PLINK 1.9 leaving 2790 individuals and 479095 variants for further analyses (exclusion of samples with call rate < 95%, selection of variants with call rate > 99% and minor allele frequency (MAF) > 1% and final samples selection using a call rate > 97%).

Kinship

Kinship and ancestry inference were calculated following UK Biobank[55] and 1M veteran program.[59] First King 2.1[60] was used to find duplicated individuals which have been recruited by two different routes. The analysis flagged 56 duplicated pairs, from which one was removed according to genotyping quality (GenomeStudio p50GC score or/and individual call rate). This leaves a set of 2734 unique individuals.

Regions of high linkage disquilibrium (LD) defined in the UK Biobank[55] were excluded from the analysis, as well as SNPs with MAF<1% or missingness >1%. King 2.1 was used to construct a relationship matrix up to 3rd degree using the King command `--kinship --degree 3` and then the function `largest_independent_vertex_set()` from the `igraph` tool http://igraph.sf.net was used to create a first set of unrelated individuals. Principal component analysis (PCA) was conducted with gcta 1.9[61] in the set of unrelated individuals with pruned SNPs using a window of 1000 markers, a step size of 80 markers and an r^2 threshold of 0.1. SNPs with large weights in PC1, PC2 or PC3 were removed, keeping at least 2/3 of the number of pruned SNPs to keep as an input of the next round of King 2.1. The second round of King 2.1 was run using the SNPs with low weights in PC1, PC2 and PC3 to avoid overestimating kinship in non-European individuals. After this round 2718 individuals were considered unrelated up to 3rd degree.

Genetic ancestry

Unrelated individuals from the 1000 Genome Project dataset were calculated using the same procedure described above, and both datasets were merged using the common SNPs. The merged genotyped data was pruned with `plink` using a window of 1000 markers a step size of 50 and a r^2 of 0.05, leaving 92K markers that were used to calculate the 20 first principal components with gcta 1.9. Ancestry for GenOMICC individuals was inferred using ADMIXTURE[62] populations defined in 1000 genomes. When one individual had a probability > 80% of pertaining to one ancestry, then the individual was assigned to this ancestry, otherwise the individual was assigned to admix ancestry as in the 1M veteran cohort.[59] According to this criterion there are 1818 individuals from European ancestry (EUR), 190 from African ancestry (AFR), 158 from East Asian ancestry (EAS), 254 from South Asian ancestry (SAS), and 301 individuals with admixed ancestry (2 or more).

Imputation

Genotype files were converted to plus strand and SNPs with Hardy-Weinberg Equilibrium (HWE) p-value$<10^{-6}$ were removed. Imputation was calculated using the TOPMed reference panel.[19] and results were given in GRCh38 human reference genome and plus strand. The imputed dataset was filtered for monogenic and low imputation quality score (r^2<0.4) using BCFtools 1.9. To perform GWAS, files in VCF format were further filtered for r^2>0.9 and converted to BGEN format using QCtools 1.3.[63]

UK Biobank imputed variants with imputation score >0.9 and overlapping our set of variants (n=5,981,137) were extracted and merged with GenOMICC data into a single BGEN file containing cases and controls using QCtools 1.3.

GWAS

Related individuals to degree 3 were removed. 13 individuals with American ancestry were removed as the sample size provided insufficient power to perform a reliable GWAS for this group. The final dataset includes 2244 individuals. Using PCA to infer genetic ancestry, there were 1676 individuals from European ancestry, 149 individuals from East Asian ancestry, 237 individuals from South Asian ancestry and 182 individuals from African ancestry (Supplementary Table 2). If age or deprivation status were missing for some individuals, the value was set to the mean of their ancestry. GWAS were performed separately for each ancestry group.

Tests for association between case-control status and allele dosage at individuals SNPs were performed by fitting logistic regression models using PLINK.[64] Independent analyses were performed for each ethnic group. All models included sex, age, mean-centered age^2, deprivation score decile of residential postcode, and the first 10 genomic principal components as covariates.

Genomic principal components were computed on the combined sample of all UK Biobank and GenOMICC participants. Specifically, 456,750 genetic variants were identified which were shared between the variants contained in the called genotypes in the GenOMICC dataset and imputed UK Biobank genotypes, which had an imputation info score above 0.95 and a minor allele frequency above 1%. After merging genotypes at these variants, variants were removed which had a minor allele frequency below 2.5%, a missingness rate above 1.5%, showed departure from Hardy-Weinberg equilibrium with a p value below 10^{-50}, or which were

within previously identified regions of high linkage disequilibrium within UK Biobank. After LD-pruning of the remaining variants to a maximum r^2 of 0.01 based on a 1000 variant window moving in 50 variants steps, using the PLINK indep-pairwise command and yielding 13,782 SNPs, the leading 20 genomic principal components were computed using FlashPCA2.[65]

GWAS results for European ancestry were filtered for MAF>0.01, HWE p-value > 10^{-50} and genotyping rate >0.99. An extra filter was added to avoid bias for using a different genotyping method and imputation panel between controls and cases. This could not be controlled for using regression because all cases and all controls were genotyped using different methods. MAF for each ancestry were compared between UK Biobank European controls and gnomAD hg38 non-Finnish Europeans downloaded in August 2020.[66] SNPs were were removed from the GWAS results following these two rules: (a) In SNPs with MAF > 10% in gnomAD, an absolute difference of 5% between gnomAD and UK biobank controls MAF (b) In SNPs with MAF <10% in gnomAD, a difference > 25% gnomAD MAF, between UK Biobank controls and gnomAD. GWAS from non-European ancestries were filtered for a MAF in UK Biobank controls corresponding to the same ancestry > 5% and then for the SNPs that passed QC in the European GWAS. To calculate differences between UK Biobank European individuals and gnomAD allele frequencies, non Finnish-Europeans gnomAD allele frequencues were used, as European UK Biobank controls are mainly non-Finnish.

Filtered GWAS for each ancestry, containing a total of ~4.7M SNPs, were combined in a trans-ethnic meta-analysis using METAL[20] standard error mode and controling for population stratification (genomic control on). Nearest genes were defined using FUMA v1.3.6 SNP2GENE function,[67] using LD R^2 > 0.6 and UK Biobank release 2 reference panel.

A sex-specific GWAS within European individuals was performed using 1180 unrelated male cases and 496 unrelated female cases and 5 UK Biobank random controls matched by sex and ancestry for each case. Test for association between case-control status and allele dosage at individual SNPs were performed by fitting a logistic regression model with PLINK. Age, mean age squared, deprivation decile of residential postcode and the first 10 principal components were added as covariates in the models.

Deprivation score The UK Data Service provides measures of deprivation based on Census Data and generated per postcode. The latest version of the Deprivation Scores were published in 2017 and are based on the 2011 census. Since only partial postcodes were available for most samples we were unable to use these indices directly. However, we generated an approximation to the scores by calculating an average weighted by population count across the top-level postcode areas.

The initial input file was part of the aggregated census data identified by DOI:10.5257/census/aggregate-2011-2.

Specifically the postcode data were downloaded from:

http://s3-eu-west-1.amazonaws.com/statistics.digitalresources.jisc.ac.uk/dkan/files/Postcode_Counts_and_Deprivation_Ranks/postcodes.zip

Population count and deprivation score for each published postcode were extracted and weighted average score calculated for each top-level postcode. We further categorised each top-level postcode score into decile and quintile bins for more coarse-grained analyses.

Whole Genome Sequencing

Whole Genome Sequencing (WGS) gVCF files were obtained for the 1667 individuals for which we had whole genome sequence data. Variants overlapping the positions of the imputed variants were called using GATk and variants with depth<8X (the minimum depth for which 95% coverage can be expected) were filtered. Individual VCF files were joined in a multi-sample VCF file for comparison with imputed variants. 1613 of these 1667 were used in the final GWAS. Samples were filtered and variants annotated using bcftools 1.9. VCF files obtained from imputation were processed in an identical manner. Alternative allele frequency was calculated with PLINK 2.0[68] for both WGS and imputed data.

Controls

UK Biobank

UK Biobank participants were considered as potential controls if they were not identified by the UK Biobank as outliers based on either genotyping missingness rate or heterogeneity, and their sex inferred from the genotypes matched their self-reported sex. For these individuals, information on sex (UKBID 31), age, ancestry, and residential postcode deprivation score decile was computed. Specifically, age was computed as age on April 1st, 2020 based on the participant's birth month (UKBID 34) and year (UKBID 52). The first part of the residential postcode of participants was computed based on the participant's home location (UKBID 22702 and 22704) and mapped to a deprivation score decile as previously described for GenOMICC participants. Ancestry was inferred as previously described for GenOMICC participants.

After excluding participants who had received PCR tests for Covid-19, based on information downloaded from the UK Biobank in August 2020, five individuals with matching inferred ancestry were sampled for each GenOMICC participant as controls. After sampling each control, individuals related up to 3rd degree were removed from the pool of potential further controls.

The 100,000 Genomes Project

Following ethical approval (14/EE/1112 and 13/EE/032), consenting participants from the 100,000 Genomes Project with a broad range of rare diseases, cancers and infection were enrolled by 13 regional NHS Genomic Medicine Centres across England and in Northern Ireland, Scotland and Wales and whole blood was drawn for DNA extraction. After quality assurance whole genome sequencing at 125 or 150 base pairs was performed by Illumina Laboratory Services on either Hiseq 2500 or Hiseq X sequencers in the Genomics England Sequencing Centre, followed by detection of small variants (single nucleotide variants and small indels) using Starling.

Test for association between cases-control status were performed by running mixed model association tests using SAIGE (v0.39). 1675 individuals from the GenOMICC study and 45,875 unrelated participants and of European ancestry were included. Genomic principal components were calculated for the combined dataset of GenOMICC participants and whole genome sequence data from the 100,000 Genomes Project. Principal Components Analysis (PCA) was performed with GCTA software using approximately 30,000 SNPs selected with minor allele frequency >0.005 and after LD pruning ($r2 < 0.1$ with a window size of 500kb). Fitting of the null logistic mixed model was performed using the SNPs used for PCA and included age, sex, squared age, age $\times$ sex and first 20 genomic principal components as covariates.

Test for association using SAIGE was performed after filtering of variants in the WGS dataset for genotype quality and minor allele frequency ≥ 0.05. GWAS-specific quality filtering was performed to include variants with minor allele count ≥ 20 for each phenotype, differential missingness between cases and controls (p-value $<1 \times 10^{-5}$) and departure from Hardy-Weinberg equilibrium (p-value $<1 \times 10^{-5}$).

Generation Scotland

Generation Scotland: Scottish Family Health Study (hereafter referred to as Generation Scotland) is a population-based cohort of 24 084 participants sampled from five regional centers across Scotland(www.generationscotland.org).[69] A large subset of participants were genotyped using either Illumina HumanOmniExpressExome-8v1_A or v1-2, and 20 032 passed QC criteria previously described.[70,71] Genotype imputation using the TOPMed reference panel was recently performed (freeze 5b) using Minimac4 v1.0 on the University of Michigan server https://imputationserver.sph.umich.edu.[72] Imputation data from 7689 unrelated (genomic sharing identical by descent estimated using PLINK1.9 < 5%) participants were used as control genotypes in a GWAS using GenOMICC cases of European ancestry, for quality check purpose of associated variants. GWAS was performed in a logistic regression framework implemented in the PLINK2 (https://www.cog-genomics.org/plink/2.0/) glm function, adjusting for age, sex and the first 10 principal components of European ancestry. These coordinates were obtained from projection to the principal components space of 1000 Genomes European population samples using KING v2.2.5[60] and a LD-pruned subset of target genotyped markers passing quality check and intersecting with the reference populations.

Validation

Clumped hits in discovery GWAS were validated using controls from Generation Scotland and 100K. To consider a hit validated, the direction of effect should be the same in all three GWAS and the p-value in both Generation Scotland and 100K had to be $p<0.05/n_{validations}$, where $n_{validations}$ is the number of significant independent loci in our analysis at the discovery threshold of $p < 5 \times 10^{-8}$.

Replication

GenOMICC EUR loci were defined usign the `clump` function of PLINK 1.9[68] and clumping parameters r2 = 0.1, pval = 5×10^{-8} and $pval_2 = 0.01$; distance to the nearest gene was calculated using ENSEMBL GRCh37 gene annotation.

No GWAS has been reported of critical illness or mortality in Covid-19. As a surrogate, to provide some replication for our findings, replication analyses were performed using Host Genetics Initiative build 37, version 2 (July 2020) B2 (hospitalised Covid-19 vs population) GWAS. Summary statistics were used from the full analysis, including all cohorts and GWAS without UK Biobank, to avoid sample overlap. Replication p-value was set to 6.25×10^{-4} (0.05/8, where 8 is the number of loci significant in the discovery).

Genome-wide meta-analysis

Meta-analysis between GenOMICC, HGI and 23andMe was performed using fixed-effect inverse variance meta-analysis in METAL,[20] with correction for genomic control on. The 23andMe study comprises cases and controls from EUR genetic ancestry group. The HGI B2 analysis is a trans-ancestry meta-analysis, with the great majority of cases being multi-ethnic European (EUR and FIN), with 238 cases of non-European ancestry (176 Admixed American, AMR, from BRACOVID study and 62 South Asian, SAS, from the GNH study).

Post-GWAS analyses

TWAS and Meta-TWAS

We performed transcriptome-wide association using the MetaXcan framework[26] and the GTEx v8 eQTL MASHR-M models available for download (http://predictdb.org/). To increase SNP coverage to perform TWAS, first GWAS summary statistics for European ancestry were imputed using the fizi[73] impute function (https://github.com/bogdanlab/fizi), 1000 genomes European population as LD reference and 30% as minimum proportion of SNPs for a region (–min-prop 0.3). Then, imputed GWAS results were harmonised, lifted over to hg38 and linked to 1000 Genomes reference panel using GWAS tools https://github.com/hakyimlab/summary-gwas-imputation/wiki/GWAS-Harmonization-And-Imputation.

Imputed and harmonised GWAS summary statistics were used to perform TWAS for whole blood and lung GTEx v8 tissues with S-PrediXcan function. Resulting p-values were corrected using the Bonferroni correction to find significant gene associations. To overcome the limitations of sample size in GTEx v8 lung and whole blood tissues, we performed a meta-twas prioritising genes with small p-values in these tissues and using GTEx v8 gene expression in all tissues and S-Multixcan.[74]

Mendelian randomisation

Two-sample summary data based Mendelian randomisation[22] was performed using the results of GenOMICC and the Genotype-Tissue expression project,[75] GTEx v7 (using SMR/HEIDI pre-prepared data: https://cnsgenomics.com/software/smr/#DataResource), with Generation Scotland[69,76] forming a linkage disequilibrium reference. GenOMICC results from those of European ancestry were used as the outcome; and GTEx (v7) whole blood expression results as the exposure. Additional data pertaining to GTEx v7 were downloaded from GTEx: https://gtexportal.org/ (accessed 20 Feb 2020, 05 Apr 2020, and 04 Jul 2020), and SMR/HEIDI from https://cnsgenomics.com/software/smr/ (accessed 03 Jul 2020). Analyses were conducted using Python 3.7.3 and SMR/HEIDI v1.03 (plots were made using SMR/HEIDI v0.711). An LD reference was created using data from the population-based Generation Scotland cohort (used with permission; described

previously[71]): from a random set of 5,000 individuals, using Plink v1.9 (www.cog-genomics.org/plink/1.9/), a set of individuals with a genomic relatedness cutoff < 0.01 was extracted; 2,778 individuals remained in the final set. All data used for the SMR/HEIDI analyses were limited to autosomal biallelic SNPs: 4,264,462 variants remained in the final merged dataset.

Significant (as per GTEx v7; nominal p-value below nominal p-value threshold) local (distance to transcriptional start site < 1Mb) eQTL from GTEx v7 whole blood for protein coding genes (as per GENCODE v19) with a MAF > 0.01 (GTEx v7 and GenOMICC) were considered as potential instrumental variables. Per variant, we first selected the Ensembl gene ID to which it was most strongly associated followed by selecting the variant to which each Ensembl gene ID was most strongly associated. Instruments were available for 4,614 unique Ensembl gene IDs.

Results were assessed based upon a list of genes selected *a priori* as of interest (Supplementary Table 4), and together as a whole. Replication of Bonferroni-corrected significant results was attempted in the results of Covid-19-Host Genetics Initiative - https://www.covid19hg.org/ - with UK Biobank excluded (July 2nd 2020 data release) using the eQTLgen expression dataset.[23] Hospitalized Covid-19 vs. population (ANA_B2_V2) was selected as the phenotype most similar to our own, and therefore the most appropriate for use as a replication cohort.

In order to further validate the analyses above, generalized summary-data Mendelian randomization (GSMR)[77] was performed using exposure data from https://www.eqtlgen.org/index.html (accessed 26/10/2020)[23] and the publicly available GenOMICC EUR data for TYK2 and IFNAR2. GSMR was performed using GCTA version 1.92.1 beta6 Linux. Pleiotropic SNPs were filtered using HEIDI-outlier test (threshold = 0.01) and instrument SNPs were selected at a genome-wide significance level ($P_{eQTL} < 5e\text{-}8$) using LD clumping (LD r^2 threshold = 0.05 and window size = 1Mb). The imputed genotypes for 50,000 unrelated individuals (based on SNP-derived genomic relatedness < 0.05 using HapMap 3 SNPs) from the UK Biobank were used as the LD reference for clumping. GSMR accounts for remaining LD not removed by LD clumping.

Genomic region plots

Genomic region plots were created using https://github.com/Geeketics/LocusZooms.

Gene-level and pathway analyses

Gene-level burden of significance in the EUR ancestry group result was calculated using MAGMA v1.08.[78] SNPs were annotated to genes by mapping based on genomic location. SNPs were assigned to a gene if the SNPs location is within 5 kb up- or down-stream of the gene region (defined as the transcription start site to transcription stop site). The MAGMA SNP-wise mean method was applied which utilises the sum of squared SNP Z-statistics as the test statistic. The 1000 Genomes Project European reference panel was used to estimate LD between SNPs.

Auxiliary files were downloaded from https://ctg.cncr.nl/software/magma on 1st September 2020. Gene location files for protein-coding genes were obtained from NCBI (ftp.ncbi.nlm.nih.gov):

`gene/DATA/GENE_INFO/Mammalia/Homo_sapiens.gene_info.gz`

on 29/04/2015, and from:

`genomes/Homo_sapiens/ARCHIVE/ANNOTATION_RELEASE.105/mapview/seq_gene.md.gz`

on 25/05/2016.

The reference data files used to estimate LD are derived from Phase 3 of the 1000 Genomes Project.

Competitive gene set enrichment analysis was conducted in MAGMA using a regression model that accounts for gene-gene correlations, to reduce bias resulting from clustering of functionally similar genes on the genome.[78] Gene sets were queried from the databases KEGG 2019, Reactome 2016, GO Biological Process 2018, Biocarta 2016 and WikiPathways 2019. The Benjamini-Hochberg procedure was used to control false discovery rate (<0.05).

Meta-analysis by information content (MAIC)

Multiple *in vitro* and *in vivo* studies have identified key host genes that either directly interact with SARS-CoV-2, or define the host response to SARS-CoV-2. We have previously conducted a systematic review of these studies.[49] In order to put the new associations from this GWAS into context, we performed a data-driven meta-analysis of gene-level results combined with pre-existing biological data using meta-analysis by information content (MAIC).[27] Briefly, MAIC combines experimental results from diverse sources in the form of ranked or unranked gene lists. The algorithm assigns a weighting to each input gene list, derived from the degree of overlap with other input lists. Each gene is then assigned a score calculated from the weightings for each gene list on which it appears. This process is repeated iteratively until all scores converge on a stable value. In order to prevent a single type of experiment from unduly biasing the results, input gene lists are assigned to categories, and a rule applied that only one weighting from each category can contribute to the score for any given gene.

Tissue and functional genomic enrichment

We downloaded the mean gene expression data summarised from RNA sequencing by the GTEx project (https://gtexportal.org/). The GTEx v7 data contain gene expressions of 19,791 genes in 48 human tissues. Gene expression values were normalized to numbers of transcripts per million reads (TPM). To measure the expression specificity of each gene in each tissue, each gene expression specificity was defined as the proportion of its expression in each tissue among all the tissues, i.e., a value ranging between 0 and 1. SNPs within the 10% most specifically expressed genes in each tissue were annotated for subsequent testing of heritability enrichment. For functional genomic enrichment analysis, we considered the inbuilt primary functional annotations v2.2 provided in the `ldsc` software (https://alkesgroup.broadinstitute.org/LDSCORE/) to annotated the SNPs.

With the annotated SNPs, we used stratified LD score regression (S-LDSC)[79] to test whether any human tissue or specific functional genomic feature is associated with sereve Covid-19. Our GWAS summary statistics were harmonized by the munge_sumstats.py procedure in `ldsc`. LD scores of HapMap3 SNPs (MHC region excluded) for gene annotations in each tissue were computed using a 1-cM window. The enrichment score was defined as the proportion of heritability captured by the annotated SNPs divided by the proportion of SNPs annotated.

Testing genetic correlations with other phenotypes

We applied both the LD score regression (LDSC)[80] and high-definition likelihood (HDL)[28] methods to evaluate the genetic correlations between Severe Covid-19 and 818 GWASed phenotypes stored on LD-Hub.[81] GWAS summary statistics were harmonized by the munge_sumstats.py procedure in the ldsc software. In the HDL analysis, we estimated the SNP-based narrow-sense heritability for each phenotype, and for the 818 complex traits GWASs, those with SNPs less than 90% overlap with the HDL reference panel were removed.

Genome build

Results are presented using Genome Reference Consortium Human Build 37. Imputed genotypes and whole-genome sequence data were lifted over from Genome Reference Consortium Human Build 38 using Picard liftoverVCF mode from GATK 4.0 which is based on the UCSC liftover tool (chain file obtained from ftp://ftp.ensembl.org/pub/assembly_mapping/homo_sapiens/GRCh38_to_GRCh37.chain.gz.[82]

Data Availability

Full summary data in support of the findings of this study are available for download from https://genomicc.org/data. Additional data can be analysed by qualified researchers in the ISARIC 4C/GenOMICC data analysis platform and through the Genomics England research environment.

The full GWAS summary statistics for the 23andMe discovery data set will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe

participants. Please visit https://research.23andMe.com/dataset-access/ for more information and to apply to access the data.

Acknowledgements

We thank the patients and their loved ones who volunteered to contribute to this study at one of the most difficult times in their lives, and the research staff in every intensive care unit who recruited patients at personal risk during the most extreme conditions we have ever witnessed in UK hospitals.

GenOMICC was funded by Sepsis Research (the Fiona Elizabeth Agnew Trust), the Intensive Care Society, a Wellcome-Beit Prize award to J. K. Baillie (Wellcome Trust 103258/Z/13/A) and a BBSRC Institute Program Support Grant to the Roslin Institute (BBS/E/D/20002172, BBS/E/D/10002070 and BBS/E/D/30002275). Whole-genome sequencing was done in partnership with Genomics England and was funded by UK Department of Health and Social Care, UKRI and LifeArc. ISARIC 4C is supported by grants from: the Medical Research Council [grant MC_PC_19059], the National Institute for Health Research (NIHR) [award CO-CIN-01] and by the NIHR Health Protection Research Unit (HPRU) in Emerging and Zoonotic Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with Liverpool School of Tropical Medicine and the University of Oxford [award 200907], NIHR HPRU in Respiratory Infections at Imperial College London with PHE [award 200927], Wellcome Trust and Department for International Development [215091/Z/18/Z], and the Bill and Melinda Gates Foundation [OPP1209135], and Liverpool Experimental Cancer Medicine Centre (Grant Reference: C18616/A25153), NIHR Biomedical Research Centre at Imperial College London [IS-BRC-1215-20013], EU Platform foR European Preparedness Against (Re-) emerging Epidemics (PREPARE) [FP7 project 602525] and NIHR Clinical Research Network for providing infrastructure support for this research. PJMO is supported by a NIHR Senior Investigator Award [award 201385]. The views expressed are those of the authors and not necessarily those of the DHSC, DID, NIHR, MRC, Wellcome Trust or PHE. HM was supported by the NIHR BRC at University College London Hospitals. The Health Research Board of Ireland (Clinical Trial Network Award 2014-12) funds collection of samples in Ireland.

This research has been conducted using the UK Biobank Resource under project 788. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award STratifying Resilience and Depression Longitudinally (STRADL) Reference 104036/Z/14/Z). Genomics England and the 100,000 Genomes Project was funded by the National Institute for Health Research, the Wellcome Trust, the Medical Research Council, Cancer Research UK, the Department of Health and Social Care and NHS England. M Caulfield is an NIHR Senior Investigator. This work is part of the portfolio of translational research at the NIHR Biomedical Research Centre at Barts and Cambridge. Research performed at the Human Genetics Unit was funded by the MRC (MC_UU_00007/10, MC_UU_00007/15). LK was supported by an RCUK Innovation Fellowship from the National Productivity Investment Fund (MR/R026408/1). A Bretherick acknowledges funding from the Wellcome Trust PhD training fellowship for clinicians (204979/Z/16/Z), and the Edinburgh Clinical Academic Track (ECAT) programme. We acknowledge support from the MRC Human Genetics Unit programme grant, "Quantitative traits in health and disease" (U. MC_UU_00007/10). A. Tenesa acknowledges funding from MRC research grant MR/P015514/1, and HDR-UK award HDR-9004 and HDR-9003.

This study owes a great deal to the National Institute of Healthcare Research Clinical Research Network (NIHR CRN) and the Chief Scientist Office (Scotland), who facilitate recruitment into research studies in NHS hospitals, and to the global ISARIC and InFACT consortia. The authors wish to thank Dr Rebecca Coll, Wellcome-Wolfson institute, Queen's Uinversity Belfast, for advice on interpretation of these results, and Dr. Jie Zheng (University of Bristol) for sharing the harmonized GWAS summary statistics used in LD-Hub. The authors wish to express their gratitude to the anonymous peer reviewers for a number of substantive improvements to the manuscript and analysis.

References

1.Carsana, L., Sonzogni, A., Nasr, A., Rossi, R.S., Pellegrinelli, A., Zerbi, P., Rech, R., Colombo, R., Antinori, S., Corbellino, M., Galli, M., Catena, E., Tosoni, A., Gianatti, A. & Nebuloni, M. Pulmonary post-mortem findings in a series of covid-19 cases from northern italy: A two-centre descriptive study. *The Lancet. Infectious diseases* (2020).doi:10.1016/S1473-3099(20)30434-5

2.Horby, P., Lim, W.S., Emberson, J.R., Mafham, M., Bell, J.L., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A., Elmahi, E., Prudon, B., Green, C., Felton, T., Chadwick, D., Rege, K., Fegan, C., Chappell, L.C., Faust, S.N., Jaki, T., Jeffery, K., Montgomery, A., Rowan, K., Juszczak, E., Baillie, J.K., Haynes, R. & Landray, M.J. Dexamethasone in hospitalized patients with covid-19 - preliminary report. *The New England journal of medicine* (2020).doi:10.1056/NEJMoa2021436

3.Baillie, J.K. Translational genomics. Targeting the host immune response to fight infection. *Science (New York, N.Y.)* **344**, 807–8(2014).

4.Dorward, D.A., Russell, C.D., Um, I.H., Elshani, M., Armstrong, S.D., Penrice-Randal, R., Millar, T., Lerpiniere, C.E., Tagliavini, G., Hartley, C.S., Randall, N.P., Gachanja, N.N., Potey, P.M., Anderson, A.M., Campbell, V.L., Duguid, A.J., Qsous, W.A., BouHaidar, R., Baillie, J.K., Dhaliwal, K., Wallace, W.A., Bellamy, C.O., Prost, S., Smith, C., Hiscox, J.A., Harrison, D.J., Lucas, C.D. & ICECAP Tissue-specific tolerance in fatal Covid-19. *medRxiv* 2020.07.02.20145003(2020).doi:10.1101/2020.07.02.20145003

5.Carvelli, J., Demaria, O., Vély, F., Batista, L., Benmansour, N.C., Fares, J., Carpentier, S., Thibult, M.-L., Morel, A., Remark, R., André, P., Represa, A., Piperoglou, C., Cordier, P.Y., Le Dault, E., Guervilly, C., Simeone, P., Gainnier, M., Morel, Y., Ebbo, M., Schleinitz, N. & Vivier, E. Association of covid-19 inflammation with activation of the c5a-c5ar1 axis. *Nature* (2020).doi:10.1038/s41586-020-2600-6

6.Sørensen, T.I., Nielsen, G.G., Andersen, P.K. & Teasdale, T.W. Genetic and environmental influences on premature death in adult adoptees. *The New England journal of medicine* **318**, 727–32(1988).

7.Patarčić, I., Gelemanović, A., Kirin, M., Kolčić, I., Theodoratou, E., Baillie, J.K., Jong, M.D. de, Rudan, I., Campbell, H. & Polašek, O. The role of host genetic factors in respiratory tract infectious diseases: Systematic review, meta-analyses and field synopsis. *Scientific reports* **5**, 16119(2015).

8.Horby, P., Nguyen, N.Y., Dunstan, S.J. & Baillie, J.K. An updated systematic review of the role of host genetics in susceptibility to influenza. *Influenza and other respiratory viruses* **7 Suppl 2**, 37–41(2013).

9.Clohisey, S. & Baillie, J.K. Host susceptibility to severe influenza A virus infection. *Critical Care* **23**, 303(2019).

10.Shelton, J.F., Shastri, A.J., Ye, C., Weldon, C.H., Filshtein-Somnez, T., Coker, D., Symons, A., Esparza-Gordillo, J., Team, T.2.C.-1., Aslibekyan, S. & Auton, A. Trans-ethnic analysis reveals genetic and non-genetic associations with COVID-19 susceptibility and severity. *medRxiv* 2020.09.04.20188318(2020).doi:10.1101/2020.09.04.20188318

11.Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., Grimsrud, M.M., Milani, C., Aziz, F., Kässens, J., May, S., Wendorff, M., Wienbrandt, L., Uellendahl-Werth, F., Zheng, T., Yi, X., Pablo, R. de, Chercoles, A.G., Palom, A., Garcia-Fernandez, A.-E., Rodriguez-Frias, F., Zanella, A., Bandera, A., Protti, A., Aghemo, A., Lleo, A., Biondi, A., Caballero-Garralda, A., Gori, A., Tanck, A., Carreras Nolla, A., Latiano, A., Fracanzani, A.L., Peschuck, A., Julià, A., Pesenti, A., Voza, A., Jiménez, D., Mateos, B., Nafria Jimenez, B., Quereda, C., Paccapelo, C., Gassner, C., Angelini, C., Cea, C., Solier, A., Pestaña, D., Muñiz-Diaz, E., Sandoval, E., Paraboschi, E.M., Navas, E., García Sánchez, F., Ceriotti, F., Martinelli-Boneschi, F., Peyvandi, F., Blasi, F., Téllez, L., Blanco-Grau, A., Hemmrich-Stanisak, G., Grasselli, G., Costantino, G., Cardamone, G., Foti, G., Aneli, S., Kurihara, H., ElAbd, H., My, I., Galván-Femenia, I., Martín, J., Erdmann, J., Ferrusquía-Acosta, J., Garcia-Etxebarria, K., Izquierdo-Sanchez, L., Bettini, L.R., Sumoy, L., Terranova, L., Moreira, L., Santoro, L., Scudeller, L., Mesonero, F., Roade, L., Rühlemann, M.C., Schaefer, M., Carrabba, M., Riveiro-Barciela, M., Figuera Basso, M.E., Valsecchi, M.G., Hernandez-Tejero, M., Acosta-Herrera, M., D'Angiò, M., Baldini, M., Cazzaniga, M., Schulzky, M., Cecconi, M., Wittig, M., Ciccarelli, M., Rodríguez-Gandía, M., Bocciolone,

M., Miozzo, M., Montano, N., Braun, N., Sacchi, N., Martínez, N., Özer, O., Palmieri, O., Faverio, P., Preatoni, P., Bonfanti, P., Omodei, P., Tentorio, P., Castro, P., Rodrigues, P.M., Blandino Ortiz, A., Cid, R. de, Ferrer, R., Gualtierotti, R., Nieto, R., Goerg, S., Badalamenti, S., Marsal, S., Matullo, G., Pelusi, S., Juzenas, S., Aliberti, S., Monzani, V., Moreno, V., Wesse, T., Lenz, T.L., Pumarola, T., Rimoldi, V., Bosari, S., Albrecht, W., Peter, W., Romero-Gómez, M., D'Amato, M., Duga, S., Banales, J.M., Hov, J.R., Folseraas, T., Valenti, L., Franke, A. & Karlsen, T.H. Genomewide association study of severe covid-19 with respiratory failure. *The New England journal of medicine* (2020).doi:10.1056/NEJMoa2020283

12.Casanova, J.-L. Severe infectious diseases of childhood as monogenic inborn errors of immunity. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E7128–37(2015).

13.Plenge, R.M. Molecular underpinnings of severe coronavirus disease 2019. *JAMA* (2020).doi:10.1001/jama.2020.14015

14.Zhang, Q., Bastard, P., Liu, Z., Pen, J.L., Moncada-Velez, M., Chen, J., Ogishi, M., Sabli, I.K.D., Hodeib, S., Korol, C., Rosain, J., Bilguvar, K., Ye, J., Bolze, A., Bigio, B., Yang, R., Arias, A.A., Zhou, Q., Zhang, Y., Onodi, F., Korniotis, S., Karpf, L., Philippot, Q., Chbihi, M., Bonnet-Madin, L., Dorgham, K., Smith, N., Schneider, W.M., Razooky, B.S., Hoffmann, H.-H., Michailidis, E., Moens, L., Han, J.E., Lorenzo, L., Bizien, L., Meade, P., Neehus, A.-L., Ugurbil, A.C., Corneau, A., Kerner, G., Zhang, P., Rapaport, F., Seeleuthner, Y., Manry, J., Masson, C., Schmitt, Y., Schlüter, A., Voyer, T.L., Khan, T., Li, J., Fellay, J., Roussel, L., Shahrooei, M., Alosaimi, M.F., Mansouri, D., Al-Saud, H., Al-Mulla, F., Almourfi, F., Al-Muhsen, S.Z., Alsohime, F., Turki, S.A., Hasanato, R., Beek, D. van de, Biondi, A., Bettini, L.R., D'Angio, M., Bonfanti, P., Imberti, L., Sottini, A., Paghera, S., Quiros-Roldan, E., Rossi, C., Oler, A.J., Tompkins, M.F., Alba, C., Vandernoot, I., Goffard, J.-C., Smits, G., Migeotte, I., Haerynck, F., Soler-Palacin, P., Martin-Nalda, A., Colobran, R., Morange, P.-E., Keles, S., Çölkesen, F., Ozcelik, T., Yasar, K.K., Senoglu, S., Karabela, Ş.N., Gallego, C.R., Novelli, G., Hraiech, S., Tandjaoui-Lambiotte, Y., Duval, X., Laouénan, C., Clinicians†, C.-S., Clinicians†, C., Group†, I.C., Group†, F.C.C.S., Cohort†, C.-C., Covid-19, A.U., Biobank†, Effort†, C.H.G., Niaid-Usuhs, Group†, T.C.I., Snow, A.L., Dalgard, C.L., Milner, J., Vinh, D.C., Mogensen, T.H., Marr, N., Spaan, A.N., Boisson, B., Boisson-Dupuis, S., Bustamante, J., Puel, A., Ciancanelli, M., Meyts, I., Maniatis, T., Soumelis, V., Amara, A., Nussenzweig, M., García-Sastre, A., Krammer, F., Pujol, A., Duffy, D., Lifton, R., Zhang, S.-Y., Gorochov, G., Béziat, V., Jouanguy, E., Sancho-Shimizu, V., Rice, C.M., Abel, L., Notarangelo, L.D., Cobat, A., Su, H.C. & Casanova, J.-L. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* (2020).doi:10.1126/science.abd4570

15.Millar, J.E., Neyton, L., Seth, S., Dunning, J., Merson, L., Murthy, S., Russell, C.D., Keating, S., Swets, M., Sudre, C.H., Spector, T.D., Ourselin, S., Steves, C.J., Wolf, J., Investigators, I., Docherty, A.B., Harrison, E.M., Openshaw, P.J., Semple, M.G. & Baillie, J.K. Robust, reproducible clinical patterns in hospitalised patients with COVID-19. *medRxiv* 2020.08.14.20168088(2020).doi:10.1101/2020.08.14.20168088

16.Docherty, A.B., Harrison, E.M., Green, C.A., Hardwick, H.E., Pius, R., Norman, L., Holden, K.A., Read, J.M., Dondelinger, F., Carson, G., Merson, L., Lee, J., Plotkin, D., Sigfrid, L., Halpin, S., Jackson, C., Gamble, C., Horby, P.W., Nguyen-Van-Tam, J.S., Ho, A., Russell, C.D., Dunning, J., Openshaw, P.J., Baillie, J.K. & Semple, M.G. Features of 200.167em133 uk patients in hospital with covid-19 using the isaric who clinical characterisation protocol: Prospective observational cohort study. *BMJ (Clinical research ed.)* **369**, m1985(2020).

17.Angus, D.C., Derde, L., Al-Beidh, F., Annane, D., Arabi, Y., Beane, A., Bentum-Puijk, W. van, Berry, L., Bhimani, Z., Bonten, M., Bradbury, C., Brunkhorst, F., Buxton, M., Buzgau, A., Cheng, A.C., Jong, M. de, Detry, M., Estcourt, L., Fitzgerald, M., Goossens, H., Green, C., Haniffa, R., Higgins, A.M., Horvat, C., Hullegie, S.J., Kruger, P., Lamontagne, F., Lawler, P.R., Linstrum, K., Litton, E., Lorenzi, E., Marshall, J., McAuley, D., McGlothin, A., McGuinness, S., McVerry, B., Montgomery, S., Mouncey, P., Murthy, S., Nichol, A., Parke, R., Parker, J., Rowan, K., Sanil, A., Santos, M., Saunders, C., Seymour, C., Turner, A., Veerdonk, F. van de, Venkatesh, B., Zarychanski, R., Berry, S., Lewis, R.J., McArthur, C., Webb, S.A. & Gordon, A.C. Effect of hydrocortisone on mortality and organ support in patients with severe covid-19: The remap-cap covid-19 corticosteroid domain randomized clinical trial. *JAMA* (2020).doi:10.1001/jama.2020.17022

18.Baillie, J.K. & Digard, P. Influenza Time to Target the Host? *New England Journal of Medicine* **369**, 191–193(2013).

19.D, T., Dn, H., Md, K., J, C., Za, S., R, T., Sag, T., A, C., Sm, G., Hm, K., An, P., J, L., S, L., X, T., Bl, B., S, D., A, E., We, C., Dp, L., Ac, S., Tw, B., Q, W., Dk, A., Ae, A.-K., Kc, B., E, B., S, G., R, G., Km, R., Ss, R., E, S., P, Q., W, G., Gj, P., Da, N., Sr, B., Mc, Z., S, Z., Jg, W., La, C., Cc, L., Ce, J., Rd, H., Td, O. & Gr, A. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. (2019).doi:10.1101/563866

20.Willer, C.J., Li, Y. & Abecasis, G.R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)* **26**, 2190–1(2010).

21.Bretherick, A.D., Canela-Xandri, O., Joshi, P.K., Clark, D.W., Rawlik, K., Boutin, T.S., Zeng, Y., Amador, C., Navarro, P., Rudan, I., Wright, A.F., Campbell, H., Vitart, V., Hayward, C., Wilson, J.F., Tenesa, A., Ponting, C.P., Baillie, J.K. & Haley, C. Linking protein to phenotype with mendelian randomization detects 38 proteins with causal roles in human diseases and traits. *PLoS genetics* **16**, e1008785(2020).

22.Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M. & Yang, J. Integration of summary data from gwas and eQTL studies predicts complex trait gene targets. *Nature genetics* **48**, 481–7(2016).

23.Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Kasela, S., Pervjakova, N., Alvaes, I., Fave, M.-J., Agbessi, M., Christiansen, M., Jansen, R., Seppälä, I., Tong, L., Teumer, A., Schramm, K., Hemani, G., Verlouw, J., Yaghootkar, H., Sönmez, R., Brown, A., Kukushkina, V., Kalnapenkis, A., Rüeger, S., Porcu, E., Kronberg-Guzman, J., Kettunen, J., Powell, J., Lee, B., Zhang, F., Arindrarto, W., Beutner, F., Consortium, B., Brugge, H., Consortium, i., Dmitreva, J., Elansary, M., Fairfax, B.P., Georges, M., Heijmans, B.T., Kähönen, M., Kim, Y., Knight, J.C., Kovacs, P., Krohn, K., Li, S., Loeffler, M., Marigorta, U.M., Mei, H., Momozawa, Y., Müller-Nurasyid, M., Nauck, M., Nivard, M., Penninx, B., Pritchard, J., Raitakari, O., Rotzchke, O., Slagboom, E.P., Stehouwer, C.D.A., Stumvoll, M., Sullivan, P., Hoen, P.A.C.∴, Thiery, J., Tönjes, A., Dongen, J. van, Iterson, M. van, Veldink, J., Völker, U., Wijmenga, C., Swertz, M., Andiappan, A., Montgomery, G.W., Ripatti, S., Perola, M., Kutalik, Z., Dermitzakis, E., Bergmann, S., Frayling, T., Meurs, J. van, Prokisch, H., Ahsan, H., Pierce, B., Lehtimäki, T., Boomsma, D., Psaty, B.M., Gharib, S.A., Awadalla, P., Milani, L., Ouwehand, W., Downes, K., Stegle, O., Battle, A., Yang, J., Visscher, P.M., Scholz, M., Gibson, G., Esko, T. & Franke, L. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* 447367(2018).doi:10.1101/447367

24.Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J. & Im, H.K. A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics* **47**, 1091–8(2015).

25.Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., Geus, E.J.C. de, Boomsma, D.I., Wright, F.A., Sullivan, P.F., Nikkola, E., Alvarez, M., Civelek, M., Lusis, A.J., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., Raitakari, O.T., Kuusisto, J., Laakso, M., Price, A.L., Pajukanta, P. & Pasaniuc, B. Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics* **48**, 245–52(2016).

26.Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., Stahl, E.A., Huckins, L.M., Nicolae, D.L., Cox, N.J. & Im, H.K. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from gwas summary statistics. *Nature communications* **9**, 1825(2018).

27.Li, B., Clohisey, S.M., Chia, B.S., Wang, B., Cui, A., Eisenhaure, T., Schweitzer, L.D., Hoover, P., Parkinson, N.J., Nachshon, A., Smith, N., Regan, T., Farr, D., Gutmann, M.U., Bukhari, S.I., Law, A., Sangesland, M., Gat-Viks, I., Digard, P., Vasudevan, S., Lingwood, D., Dockrell, D.H., Doench, J.G., Baillie, J.K. & Hacohen, N. Genome-wide CRISPR screen identifies host dependency factors for influenza A virus infection. *Nature Communications* **11**, 164(2020).

28.Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations across human complex traits. *Nature genetics* **52**, 859–864(2020).

29.Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J.,

Lilje, B., Rapin, N., Bagger, F.O., Jørgensen, M., Andersen, P.R., Bertin, N., Rackham, O., Burroughs, A.M., Baillie, J.K., Ishizu, Y., Shimizu, Y., Furuhata, E., Maeda, S., Negishi, Y., Mungall, C.J., Meehan, T.F., Lassmann, T., Itoh, M., Kawaji, H., Kondo, N., Kawai, J., Lennartsson, A., Daub, C.O., Heutink, P., Hume, D.A., Jensen, T.H., Suzuki, H., Hayashizaki, Y., Müller, F., Consortium, T.F., Forrest, A.R.R., Carninci, P., Rehli, M. & Sandelin, A. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461(2014).

30.Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., Turner, J.M.A., Bertelsen, M.F., Murchison, E.P., Flicek, P. & Odom, D.T. Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–66(2015).

31.Plenge, R.M., Scolnick, E.M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nature Reviews Drug Discovery* **12**, 581–594(2013).

32.Davies, N.M., Holmes, M.V. & Smith, G.D. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ* **362**, k601(2018).

33.Sadler, A.J. & Williams, B.R.G. Interferon-inducible antiviral effectors. *Nature reviews. Immunology* **8**, 559–68(2008).

34.Duncan, C.J.A., Mohamad, S.M.B., Young, D.F., Skelton, A.J., Leahy, T.R., Munday, D.C., Butler, K.M., Morfopoulou, S., Brown, J.R., Hubank, M., Connell, J., Gavin, P.J., McMahon, C., Dempsey, E., Lynch, N.E., Jacques, T.S., Valappil, M., Cant, A.J., Breuer, J., Engelhardt, K.R., Randall, R.E. & Hambleton, S. Human ifnar2 deficiency: Lessons for antiviral immunity. *Science translational medicine* **7**, 307ra154(2015).

35.Hambleton, S., Goodbourn, S., Young, D.F., Dickinson, P., Mohamad, S.M.B., Valappil, M., McGovern, N., Cant, A.J., Hackett, S.J., Ghazal, P., Morgan, N.V. & Randall, R.E. STAT2 deficiency and susceptibility to viral illness in humans. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 3053–8(2013).

36.Consortium, W.S. trial, Pan, H., Peto, R., Karim, Q.A., Alejandria, M., Henao-Restrepo, A.M., García, C.H., Kieny, M.-P., Malekzadeh, R., Murthy, S., Preziosi, M.-P., Reddy, S., Periago, M.R., Sathiyamoorthy, V., Røttingen, J.-A., Swaminathan, S. & as the members of the Writing Committee, A.R. for the C. and I. of T.A. Repurposed antiviral drugs for COVID-19 interim WHO SOLIDARITY trial results. *medRxiv* 2020.10.15.20209817(2020).doi:10.1101/2020.10.15.20209817

37.Hamano, E., Hijikata, M., Itoyama, S., Quy, T., Phi, N.C., Long, H.T., Ha, L.D., Ban, V.V., Matsushita, I., Yanai, H., Kirikae, F., Kirikae, T., Kuratsuji, T., Sasazuki, T. & Keicho, N. Polymorphisms of interferon-inducible genes oas-1 and mxa associated with sars in the vietnamese population. *Biochemical and biophysical research communications* **329**, 1234–9(2005).

38.He, J., Feng, D., Vlas, S.J. de, Wang, H., Fontanet, A., Zhang, P., Plancoulaine, S., Tang, F., Zhan, L., Yang, H., Wang, T., Richardus, J.H., Habbema, J.D.F. & Cao, W. Association of sars susceptibility with single nucleic acid polymorphisms of oas1 and mxa genes: A case-control study. *BMC infectious diseases* **6**, 106(2006).

39.Choi, U.Y., Kang, J.-S., Hwang, Y.S. & Kim, Y.-J. Oligoadenylate synthase-like (oasl) proteins: Dual functions and associations with diseases. *Experimental & molecular medicine* **47**, e144(2015).

40.Hagemeijer, M.C., Vonk, A.M., Monastyrska, I., Rottier, P.J.M. & Haan, C.A.M. de Visualizing coronavirus rna synthesis in time by using click chemistry. *Journal of virology* **86**, 5808–16(2012).

41.Wood, E.R., Bledsoe, R., Chai, J., Daka, P., Deng, H., Ding, Y., Harris-Gurley, S., Kryn, L.H., Nartey, E., Nichols, J., Nolte, R.T., Prabhu, N., Rise, C., Sheahan, T., Shotwell, J.B., Smith, D., Tai, V., Taylor, J.D., Tomberlin, G., Wang, L., Wisely, B., You, S., Xia, B. & Dickson, H. The Role of Phosphodiesterase 12 (PDE12) as a Negative Regulator of the Innate Immune Response and the Discovery of Antiviral Inhibitors. *Journal of Biological Chemistry* jbc.M115.653113(2015).doi:10.1074/jbc.M115.653113

42.Silverman, R.H. & Weiss, S.R. Viral Phosphodiesterases That Antagonize Double-Stranded RNA Signaling to RNase L by Degrading 2-5A. *Journal of Interferon & Cytokine Research* **34**, 455–463(2014).

43.Fingerlin, T.E., Murphy, E., Zhang, W., Peljto, A.L., Brown, K.K., Steele, M.P., Loyd, J.E., Cosgrove, G.P., Lynch, D., Groshong, S., Collard, H.R., Wolters, P.J., Bradford, W.Z., Kossen, K., Seiwert, S.D., Bois, R.M. du, Garcia, C.K., Devine, M.S., Gudmundsson, G., Isaksson, H.J., Kaminski, N., Zhang, Y., Gibson, K.F., Lancaster, L.H., Cogan, J.D., Mason, W.R., Maher, T.M., Molyneaux, P.L., Wells, A.U., Moffatt, M.F., Selman, M., Pardo, A., Kim, D.S., Crapo, J.D., Make, B.J., Regan, E.A., Walek, D.S., Daniel, J.J., Kamatani, Y., Zelenika, D., Smith, K., McKean, D., Pedersen, B.S., Talbert, J., Kidd, R.N., Markin, C.R., Beckman, K.B., Lathrop, M., Schwarz, M.I. & Schwartz, D.A. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nature genetics* **45**, 613–20(2013).

44.Allen, R.J., Guillen-Guio, B., Oldham, J.M., Ma, S.-F., Dressen, A., Paynton, M.L., Kraven, L.M., Obeidat, M., Li, X., Ng, M., Braybrooke, R., Molina-Molina, M., Hobbs, B.D., Putman, R.K., Sakornsakolpat, P., Booth, H.L., Fahy, W.A., Hart, S.P., Hill, M.R., Hirani, N., Hubbard, R.B., McAnulty, R.J., Millar, A.B., Navaratnam, V., Oballa, E., Parfrey, H., Saini, G., Whyte, M.K.B., Zhang, Y., Kaminski, N., Adegunsoye, A., Strek, M.E., Neighbors, M., Sheng, X.R., Gudmundsson, G., Gudnason, V., Hatabu, H., Lederer, D.J., Manichaikul, A., Newell, J.D., O'Connor, G.T., Ortega, V.E., Xu, H., Fingerlin, T.E., Bossé, Y., Hao, K., Joubert, P., Nickle, D.C., Sin, D.D., Timens, W., Furniss, D., Morris, A.P., Zondervan, K.T., Hall, I.P., Sayers, I., Tobin, M.D., Maher, T.M., Cho, M.H., Hunninghake, G.M., Schwartz, D.A., Yaspan, B.L., Molyneaux, P.L., Flores, C., Noth, I., Jenkins, R.G. & Wain, L.V. Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine* **201**, 564–574(2020).

45.Zhang, H., Maqsudi, S., Rainczuk, A., Duffield, N., Lawrence, J., Keane, F.M., Justa-Schuch, D., Geiss-Friedlander, R., Gorrell, M.D. & Stephens, A.N. Identification of novel dipeptidyl peptidase 9 substrates by two-dimensional differential in-gel electrophoresis. *The FEBS journal* **282**, 3737–57(2015).

46.Geiss-Friedlander, R., Parmentier, N., Möller, U., Urlaub, H., Van den Eynde, B.J. & Melchior, F. The cytoplasmic peptidase dpp9 is rate-limiting for degradation of proline-containing peptides. *The Journal of biological chemistry* **284**, 27211–9(2009).

47.Griswold, A.R., Ball, D.P., Bhattacharjee, A., Chui, A.J., Rao, S.D., Taabazuing, C.Y. & Bachovchin, D.A. DPP9's enzymatic activity and not its binding to card8 inhibits inflammasome activation. *ACS chemical biology* **14**, 2424–2429(2019).

48.Nguyen, D.-T., Mathias, S., Bologa, C., Brunak, S., Fernandez, N., Gaulton, A., Hersey, A., Holmes, J., Jensen, L.J., Karlsson, A., Liu, G., Ma'ayan, A., Mandava, G., Mani, S., Mehta, S., Overington, J., Patel, J., Rouillard, A.D., Schürer, S., Sheils, T., Simeonov, A., Sklar, L.A., Southall, N., Ursu, O., Vidovic, D., Waller, A., Yang, J., Jadhav, A., Oprea, T.I. & Guha, R. Pharos: Collating protein information to shed light on the druggable genome. *Nucleic acids research* **45**, D995–D1002(2017).

49.Parkinson, N., Rodgers, N., Fourman, M.H., Wang, B., Zechner, M., Swets, M., Millar, J.E., Law, A., Russell, C., Baillie, J.K. & Clohisey, S. Systematic review and meta-analysis identifies potential host therapeutic targets in COVID-19. *medRxiv* 2020.08.27.20182238(2020).doi:10.1101/2020.08.27.20182238

50.Zhou, Z., Ren, L., Zhang, L., Zhong, J., Xiao, Y., Jia, Z., Guo, L., Yang, J., Wang, C., Jiang, S., Yang, D., Zhang, G., Li, H., Chen, F., Xu, Y., Chen, M., Gao, Z., Yang, J., Dong, J., Liu, B., Zhang, X., Wang, W., He, K., Jin, Q., Li, M. & Wang, J. Heightened innate immune responses in the respiratory tract of covid-19 patients. *Cell host & microbe* **27**, 883–890.e2(2020).

51.Zhao, Y., Qin, L., Zhang, P., Li, K., Liang, L., Sun, J., Xu, B., Dai, Y., Li, X., Zhang, C., Peng, Y., Feng, Y., Li, A., Hu, Z., Xiang, H., Ogg, G., Ho, L.-P., McMichael, A., Jin, R., Knight, J.C., Dong, T. & Zhang, Y. Longitudinal covid-19 profiling associates il-1RA and il-10 with disease severity and rantes with mild disease. *JCI insight* **5**, (2020).

52.Vergunst, C.E., Gerlag, D.M., Lopatinskaya, L., Klareskog, L., Smith, M.D., Bosch, F. van den, Dinant, H.J., Lee, Y., Wyant, T., Jacobson, E.W., Baeten, D. & Tak, P.P. Modulation of ccr2 in rheumatoid arthritis: A double-blind, randomized, placebo-controlled clinical trial. *Arthritis and rheumatism* **58**, 1931–9(2008).

53.Docherty, A.B., Harrison, E.M., Green, C.A., Hardwick, H.E., Pius, R., Norman, L., Holden, K.A., Read, J.M., Dondelinger, F., Carson, G., Merson, L., Lee, J., Plotkin, D., Sigfrid, L., Halpin, S., Jackson, C.,

Gamble, C., Horby, P.W., Nguyen-Van-Tam, J.S., Ho, A., Russell, C.D., Dunning, J., Openshaw, P.J., Baillie, J.K. & Semple, M.G. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Prospective observational cohort study. *BMJ* **369**, (2020).

54.Williamson, E.J., Walker, A.J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C.E., Curtis, H.J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H.I., MacKenna, B., Tomlinson, L., Douglas, I.J., Rentsch, C.T., Mathur, R., Wong, A.Y.S., Grieve, R., Harrison, D., Forbes, H., Schultze, A., Croker, R., Parry, J., Hester, F., Harper, S., Perera, R., Evans, S.J.W., Smeeth, L. & Goldacre, B. Factors associated with covid-19-related death using opensafely. *Nature* **584**, 430–436(2020).

55.Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T. & Collins, R. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**, e1001779(2015).

56.McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M.A. The genome analysis toolkit: A mapreduce framework for analyzing next-generation dna sequencing data. *Genome research* **20**, 1297–303(2010).

57.Meynert, A.M., Ansari, M., FitzPatrick, D.R. & Taylor, M.S. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC bioinformatics* **15**, 247(2014).

58.Guo, Y., He, J., Zhao, S., Wu, H., Zhong, X., Sheng, Q., Samuels, D.C., Shyr, Y. & Long, J. Illumina human exome genotyping array clustering and quality control. *Nature protocols* **9**, 2643–62(2014).

59.Gaziano, J.M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., Guarino, P., Aslan, M., Anderson, D., LaFleur, R., Hammond, T., Schaa, K., Moser, J., Huang, G., Muralidhar, S., Przygodzki, R. & O'Leary, T.J. Million veteran program: A mega-biobank to study genetic influences on health and disease. *Journal of clinical epidemiology* **70**, 214–23(2016).

60.Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M. & Chen, W.-M. Robust relationship inference in genome-wide association studies. *Bioinformatics (Oxford, England)* **26**, 2867–73(2010).

61.Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76–82(2011).

62.Alexander, D.H. & Lange, K. Enhancements to the admixture algorithm for individual ancestry estimation. *BMC bioinformatics* **12**, 246(2011).

63.Wigginton, J.E., Cutler, D.J. & Abecasis, G.R. A note on exact tests of hardy-weinberg equilibrium. *American journal of human genetics* **76**, 887–93(2005).

64.Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. Second-generation plink: Rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7(2015).

65.Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: Principal component analysis of biobank-scale genotype datasets. *Bioinformatics (Oxford, England)* **33**, 2776–2778(2017).

66.Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Neale, B.M., Daly, M.J. & MacArthur, D.G. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443(2020).

67.Watanabe, K., Taskesen, E., Bochoven, A. van & Posthuma, D. Functional mapping and annotation of genetic associations with fuma. *Nature communications* **8**, 1826(2017).

68.Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., Bakker, P.I.W. de, Daly, M.J. & Sham, P.C. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559–75(2007).

69.Smith, B.H., Campbell, A., Linksted, P., Fitzpatrick, B., Jackson, C., Kerr, S.M., Deary, I.J., Macintyre, D.J., Campbell, H., McGilchrist, M., Hocking, L.J., Wisely, L., Ford, I., Lindsay, R.S., Morton, R., Palmer, C.N.A., Dominiczak, A.F., Porteous, D.J. & Morris, A.D. Cohort profile: Generation scotland: Scottish family health study (gs:SFHS). The study, its participants and their potential for genetic research on health and illness. *International journal of epidemiology* **42**, 689–700(2013).

70.Amador, C., Huffman, J., Trochet, H., Campbell, A., Porteous, D., Wilson, J.F., Hastie, N., Vitart, V., Hayward, C., Navarro, P. & Haley, C.S. Recent genomic heritage in scotland. *BMC genomics* **16**, 437(2015).

71.Nagy, R., Boutin, T.S., Marten, J., Huffman, J.E., Kerr, S.M., Campbell, A., Evenden, L., Gibson, J., Amador, C., Howard, D.M., Navarro, P., Morris, A., Deary, I.J., Hocking, L.J., Padmanabhan, S., Smith, B.H., Joshi, P., Wilson, J.F., Hastie, N.D., Wright, A.F., McIntosh, A.M., Porteous, D.J., Haley, C.S., Vitart, V. & Hayward, C. Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 generation scotland participants. *Genome medicine* **9**, 23(2017).

72.Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.-R., Iacono, W.G., Swaroop, A., Scott, L.J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G.R. & Fuchsberger, C. Next-generation genotype imputation service and methods. *Nature genetics* **48**, 1284–1287(2016).

73.Pasaniuc, B., Zaitlen, N., Shi, H., Bhatia, G., Gusev, A., Pickrell, J., Hirschhorn, J., Strachan, D.P., Patterson, N. & Price, A.L. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics (Oxford, England)* **30**, 2906–14(2014).

74.Barbeira, A.N., Pividori, M., Zheng, J., Wheeler, H.E., Nicolae, D.L. & Im, H.K. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS genetics* **15**, e1007889(2019).

75.Battle, A., Brown, C.D., Engelhardt, B.E. & Montgomery, S.B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213(2017).

76.Smith, B.H., Campbell, H., Blackwood, D., Connell, J., Connor, M., Deary, I.J., Dominiczak, A.F., Fitzpatrick, B., Ford, I., Jackson, C., Haddow, G., Kerr, S., Lindsay, R., McGilchrist, M., Morton, R., Murray, G., Palmer, C.N.A., Pell, J.P., Ralston, S.H., St Clair, D., Sullivan, F., Watt, G., Wolf, R., Wright, A., Porteous, D. & Morris, A.D. Generation Scotland: The Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Medical Genetics* **7**, 74(2006).

77.Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M.R., McGrath, J.J., Visscher, P.M., Wray, N.R. & Yang, J. Causal associations between risk factors and common diseases inferred from gwas summary data. *Nature communications* **9**, 224(2018).

78.Leeuw, C.A. de, Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: Generalized gene-set analysis of gwas data. *PLoS computational biology* **11**, e1004219(2015).

79.Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F.R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J.R.B., Okada, Y., Raychaudhuri, S., Daly, M.J., Patterson, N., Neale, B.M. & Price, A.L. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228–35(2015).

80.Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., Daly, M.J., Price, A.L. & Neale, B.M. An atlas of genetic correlations across human diseases and traits. *Nature genetics* **47**, 1236–41(2015).

81.Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B.S., Warrington, N.M., Finucane, H.K., Price, A.L., Bulik-Sullivan, B.K., Anttila, V., Paternoster, L., Gaunt, T.R., Evans, D.M. & Neale, B.M. LD hub: A centralized database and web interface to perform ld score regression that maximizes the potential of summary level gwas data for snp heritability and genetic correlation analysis. *Bioinformatics (Oxford, England)* **33**, 272–279(2017).

82.Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., Hillman-Jackson, J., Kuhn, R.M., Pedersen, J.S., Pohl, A., Raney, B.J., Rosenbloom, K.R., Siepel, A., Smith, K.E., Sugnet, C.W., Sultan-Qurraie, A., Thomas, D.J., Trumbower, H., Weber, R.J., Weirauch, M., Zweig, A.S., Haussler, D. & Kent, W.J. The ucsc genome browser database: Update 2006. *Nucleic acids research* **34**, D590–8(2006).